Visually searching the Web for content

- *Smith, J.R.; Shih-Fu Chang*
Columbia Univ., NY, USA

This paper appears in: **IEEE Multimedia**
On page(s): 12 - 20
July-Sept. 1997
Volume: 4 Issue: 3
ISSN: 1070-986X
References Cited: 11
CODEN: IEMUE4
INSPEC Accession Number: 5705228

Abstract:
New visual information in the form of images, graphics, animations and videos is published on the World Wide Web at an incredible rate. However, cataloging it exceeds the capabilities of current text-based Web search engines. WebSeek provides a complet system that collects visual information from the Web by automated agents, then catalog and indexes it for fast searching and retrieval.

Index Terms:
Internet; visual searching; World Wide Web content searching; visual information; images; graphics; animation; videos; publishing; cataloging; search engines; WebSeek automated agents; indexing; information retrieval

Customized metadata for Internet information

- *Dunkel, B.; Soparkar, N.; Weinstein, P.*
Editor(s): Jain, L.C.
Dept. of Electr. Eng. & Comput. Sci., Michigan Univ., Ann Arbor, MI, USA

This paper appears in: **Knowledge-Based Intelligent Electronic Systems, 1997. KE '97. Proceedings., 1997 First International Conference on**
On page(s): 508 - 516 vol.2
21-23 May 1997
1997
Volume: 2
ISBN: 0-7803-3755-7
IEEE Catalog Number: 97TH8250
Number of Pages: 2 vol. 697
References Cited: 11
INSPEC Accession Number: 5704780

Abstract:
Several search engines, catalogs, and filtering services aim to help users of the Internet deal with a growing information "overload". However, these tools typically are either generic in scope, or limited to the needs of a particular user without regard for reuse in some related context. We propose an approach and architecture for customized filtering and cataloging which bridges these two extremes. We allow users to create and maintai a metadatabase of information gathered over time by using modular filters created for specific needs or drawn from a standard library. This metadatabase, which may be regarded as a database view of the Internet, can then be accessed to locate information relevant to specific or more generic tasks. Potentially, our approach achieves greater flexibility and specificity as compared to currently available tools. We describe our preliminary design, implementation, and experimentation for our proof of concept prototypical effort.

Index Terms:
Internet; customized metadata; Internet information; search engines; catalogs; filtering services; Internet users; customized filtering; metadatabase; modular filters; standard library; database view; proof of concept prototypical effort

WISE: a World Wide Web resource database system

- *Budi Yuwono; Dik Lun Lee*
Dept. of Comput. & Inf. Sci., Ohio State Univ., Columbus, OH, USA

This paper appears in: **Knowledge and Data Engineering, IEEE Transactions on**
On page(s): 548 - 554
Aug. 1996
Volume: 8 Issue: 4
ISSN: 1041-4347
References Cited: 17
INSPEC Accession Number: 5378930

Abstract:
The paper describes the World Wide Web Index and Search Engine (WISE) for Internet resource discovery. The system is designed around a resource database containing meta information about WWW resources and is automatically built using an indexer robot, a special WWW client agent. The resource database allows users to search for resources based on keywords, and to learn about potentially relevant resources without having to directly access them. Such capabilities can significantly reduce the amount of time that user needs to spend in order to find the information of his/her interest. We discuss WIS main components: the resource database, the indexer robot, the search engine, and the user interface, and through the technical discussions, we highlight the research issues involved in the design, the implementation and the evaluation of such a system.

Index Terms:
Internet; information retrieval; online front-ends; indexing; database management systems; user interfaces; WISE; World Wide Web resource database system; World Wid Web Index and Search Engine; Internet resource discovery; meta information; WWW resources; indexer robot; special WWW client agent; search engine; user interface

Your search matched **298** of **676793** documents.
Results are shown **25** to a page, sorted by **publication year** in **descending** order.
You may refine your search by editing the current search expression or entering a new one the text box.
Then click **Search Again**.

search engines

Search Again

Results:
Journal or Magazine = **JNL** Conference = **CNF** Standard = **STD**

251 **Distinguishing photographs and graphics on the World Wide Web**
Athitsos, V.; Swain, M.J.; Frankel, C.
Content-Based Access of Image and Video Libraries, 1997. Proceedings IEEE Worksho on , 1997
Page(s): 10 -17

[Abstract] [PDF Full-Text] **CNF**

252 **ImageRover: a content-based image browser for the World Wide Web**
Sclaroff, S.; Taycher, L.; La Cascia, M.
Content-Based Access of Image and Video Libraries, 1997. Proceedings IEEE Worksho on , 1997
Page(s): 2 -9

[Abstract] [PDF Full-Text] **CNF**

253 **Villa: an event based middleware for real-time collaboration on the Internet**
Mukherjee, B.; Bhola, S.; Doddapaneni, S.
Community Networking Proceedings, 1997 Fourth International Workshop on , 1997
Page(s): 65 -74

[Abstract] [PDF Full-Text] **CNF**

254 **Indexing Byzantine Cypriot coinage, using image analysis**
Lanitis, A.; Neophytou, A.
Digital Signal Processing Proceedings, 1997. DSP 97., 1997 13th International Conference on , Volume: 2 , 1997
Page(s): 1153 -1156 vol.2

[Abstract] [PDF Full-Text] **CNF**

255 **Visualising sequences of queries: a new tool for information retrieval**
Beale, R.; McNab, R.J.; Witten, I.H.
Information Visualization, 1997. Proceedings., 1997 IEEE Conference on , 1997
Page(s): 57 -62

[Abstract] [PDF Full-Text] **CNF**

256 **A natural language processing based Internet agent**
Ming-Hsuan Yang; Yang, C.C.; Yi-Ming Chung
Systems, Man, and Cybernetics, 1997. Computational Cybernetics and Simulation., 1997 IEEE International Conference on , Volume: 1 , 1997
Page(s): 100 -105 vol.1

[Abstract] [PDF Full-Text] **CNF**

257 **A content-based search engine on medical images for telemedicine**
Cheung, D.; Chi-Hung Lee; Ng, V.
Computer Software and Applications Conference, 1997. COMPSAC '97. Proceedings., The Twenty-First Annual International , 1997
Page(s): 569 -572

[Abstract] [PDF Full-Text] **CNF**

258 **RMESH algorithms for parallel string matching**
Hsi-Chieh Lee; Ercal, F.
Parallel Architectures, Algorithms, and Networks, 1997. (I-SPAN '97) Proceedings., Third International Symposium on , 1997
Page(s): 223 -226

[Abstract] [PDF Full-Text] **CNF**

259 **A relational interface for heterogeneous information sources**
Ee-Peng Lim; Wee Keong Ng
Digital Libraries, 1997. ADL '97. Proceedings., IEEE International Forum on Research and Technology Advances in , 1997
Page(s): 128 -139

[Abstract] [PDF Full-Text] **CNF**

260 **Effectiveness of cooperative resource collecting robots for Web search engines**
Kamei, S.; Kawano, H.; Hasegawa, T.
Communications, Computers and Signal Processing, 1997. 10 Years PACRIM 1987-19 - Networking the Pacific Rim. 1997 IEEE Pacific Rim Conference on , Volume: 1 , 1997
Page(s): 410 -413 vol.1

[Abstract] [PDF Full-Text] **CNF**

261 **Mondou: Web search engine with textual data mining**
Kawano, H.
Communications, Computers and Signal Processing, 1997. 10 Years PACRIM 1987-19 - Networking the Pacific Rim. 1997 IEEE Pacific Rim Conference on , Volume: 1 , 1997
Page(s): 402 -405 vol.1

[Abstract] [PDF Full-Text] **CNF**

262 **On using the Web as infrastructure for reengineering**

Tilley, S.R.; Smith, D.B.
Program Comprehension, 1997. IWPC '97. Proceedings., Fifth Iternational Workshop o , 1997
Page(s): 170 -173

[Abstract] [PDF Full-Text] **CNF**

---

263 **Catalytic inference analysis: detecting inference threats due to knowledg discovery**
Hale, J.; Shenoi, S.
Security and Privacy, 1997. Proceedings., 1997 IEEE Symposium on , 1997
Page(s): 188 -199

[Abstract] [PDF Full-Text] **CNF**

---

264 **An inter/intranet multimedia service for telemedicine**
Bella, R.D.; Iovieno, L.; Marcocci, M.; Masciocchi, C.; Patricelli, F.
EUROMICRO 97. New Frontiers of Information Technology., Proceedings of the 23rd EUROMICRO Conference , 1997
Page(s): 379 -386

[Abstract] [PDF Full-Text] **CNF**

---

265 **Information retrieval and database architecture for conventional Japanes character dictionaries**
Schmitt, L.M.; Herder, J.; Bhalla, S.
Cognitive Technology, 1997. Humanizing the Information Age. Proceedings., Second International Conference on , 1997
Page(s): 200 -217

[Abstract] [PDF Full-Text] **CNF**

---

266 **Customized metadata for Internet information**
Dunkel, B.; Soparkar, N.; Weinstein, P.
Knowledge-Based Intelligent Electronic Systems, 1997. KES '97. Proceedings., 1997 First International Conference on , Volume: 2 , 1997
Page(s): 508 -516 vol.2

[Abstract] [PDF Full-Text] **CNF**

---

267 **A client-oriented distribution architecture for Web search agents**
Hyunsuk Seung; Doo-Hwan Bae
Autonomous Decentralized Systems, 1997. Proceedings. ISADS 97., Third Internation Symposium on , 1997
Page(s): 299 -308

[Abstract] [PDF Full-Text] **CNF**

---

268 **Personal Adaptive Web agent: a tool for information filtering**
Khan, I.; Card, H.C.
Electrical and Computer Engineering, 1997. Engineering Innovation: Voyage of Discovery. IEEE 1997 Canadian Conference on , Volume: 1 , 1997

Page(s): 305 -308 vol.1

[Abstract] [PDF Full-Text] **CNF**

---

269 **Visualization of satellite derived time-series datasets using computer graphics and computer animation**
Meisner, R.E.; Bittner, M.; Dech, S.W.
Geoscience and Remote Sensing, 1997. IGARSS '97. Remote Sensing - A Scientific Vision for Sustainable Development., 1997 IEEE International , Volume: 4 , 1997
Page(s): 1495 -1498 vol.4

[Abstract] [PDF Full-Text] **CNF**

---

270 **Enhancing image search engines in visual information environments**
Smith, J.R.; Shih-Fu Chang
Multimedia Signal Processing, 1997., IEEE First Workshop on , 1997
Page(s): 545 -552

[Abstract] [PDF Full-Text] **CNF**

---

271 **Adding full text indexing to the operating system**
Peltonen, K.
Data Engineering, 1997. Proceedings. 13th International Conference on , 1997
Page(s): 386 -390

[Abstract] [PDF Full-Text] **CNF**

---

272 **A real-time matching system for large fingerprint databases**
Ratha, N.K.; Karu, K.; Shaoyun Chen; Jain, A.K.
Pattern Analysis and Machine Intelligence, IEEE Transactions on , Volume: 18 Issue: 8 Aug. 1996
Page(s): 799 -813

[Abstract] [PDF Full-Text] **JNL**

---

273 **WISE: a World Wide Web resource database system**
Budi Yuwono; Dik Lun Lee
Knowledge and Data Engineering, IEEE Transactions on , Volume: 8 Issue: 4 , Aug. 1996
Page(s): 548 -554

[Abstract] [PDF Full-Text] **JNL**

---

274 **Indexes are "out", models are "in"**
Davenport, G.
IEEE Multimedia , Volume: 3 Issue: 3 , Fall 1996
Page(s): 10 -15

[Abstract] [PDF Full-Text] **JNL**

---

275 **Visualizing the results of multimedia Web search engines**

Mukherjea, S.; Hirata, K.; Hara, Y.
Information Visualization '96, Proceedings IEEE Symposium on , 1996
Page(s): 64 -65, 122

[Abstract] [PDF Full-Text] **CNF**

[Prev] 1 2 3 4 5 6 7 8 9 10 11 12 [Next]

Home | Log-out | Journals | Conference Proceedings | Standards
Search by Author | Basic Search | Advanced Search | Join IEEE | Establish a Web Account

Your search matched **3** of **676793** documents.
Results are shown **25** to a page, sorted by **publication year** in **descending** order.
You may refine your search by editing the current search expression or entering a new one the text box.
Then click **Search Again**.

meta search engines

Search Again

Results:
Journal or Magazine = **JNL** Conference = **CNF** Standard = **STD**

1 **Concept hierarchy based text database categorization in a metasearch engine environment**
Wang, W.; Meng, W.; Yu, C.
Web Information Systems Engineering, 2000. Proceedings of the First International Conference on , Volume: 1 , 2000
Page(s): 283 -290 vol.1

[Abstract] [PDF Full-Text] **CNF**

2 **SOLO: an MPEG-7 optimum search tool**
Lay, J.A.; Ling Guan
Multimedia and Expo, 2000. ICME 2000. 2000 IEEE International Conference on , Volume: 2 , 2000
Page(s): 777 -780 vol.2

[Abstract] [PDF Full-Text] **CNF**

3 **A Boolean query processing with a result cache in mediator systems**
Jae-Heon Cheong; Sang-Goo Lee
Advances in Digital Libraries, 2000. Proceedings. IEEE , 2000
Page(s): 218 -227

[Abstract] [PDF Full-Text] **CNF**

Your search matched **298** of **676793** documents.
Results are shown **25** to a page, sorted by **publication year** in **descending** order.
You may refine your search by editing the current search expression or entering a new one the text box.
Then click **Search Again**.

search engines

Search Again

Results:
Journal or Magazine = **JNL** Conference = **CNF** Standard = **STD**

276 **IEEE GLOBECOM 1996. Communications: The Key to Global Prosperity. GLOBAL INTERNET'96. Conference Record**
Global Telecommunications Conference, 1996. GLOBECOM '96. 'Communications: The Key to Global Prosperity , 1996
[Abstract] [PDF Full-Text] **CNF**

277 **Surfin' network resources across the Web**
Deri, L.
Systems Management, 1996., Proceedings of Second IEEE International Workshop on 1996
Page(s): 158 -167

[Abstract] [PDF Full-Text] **CNF**

278 **A 14 GOPS programmable motion estimator for H.26x video coding**
Horng-Dar Lin; Anesko, A.; Petryna, B.
Solid-State Circuits Conference, 1996. Digest of Technical Papers. 42nd ISSCC., 1996 IEEE International , 1996
Page(s): 246, 454
[Abstract] [PDF Full-Text] **CNF**

279 **Using the functional data model to integrate distributed biological data sources**
Kemp, G.J.L.; Dupont, J.; Gray, P.M.D.
Scientific and Statistical Database Systems, 1996. Proceedings., Eighth International Conference on , 1996
Page(s): 176 -185

[Abstract] [PDF Full-Text] **CNF**

280 **A real-time matching system for large fingerprint databases**
Ratha, N.K.; Karu, K.; Shaoyun Chen; Jain, A.K.
Pattern Analysis and Machine Intelligence, IEEE Transactions on , Volume: 18 Issue: 8 Aug. 1996
Page(s): 799 -813

[Abstract] [PDF Full-Text] **JNL**

281 **WISE: a World Wide Web resource database system**
Budi Yuwono; Dik Lun Lee
Knowledge and Data Engineering, IEEE Transactions on , Volume: 8 Issue: 4 , Aug. 1996
Page(s): 548 -554

[Abstract] [PDF Full-Text] **JNL**

---

282 **Indexes are "out", models are "in"**
Davenport, G.
IEEE Multimedia , Volume: 3 Issue: 3 , Fall 1996
Page(s): 10 -15

[Abstract] [PDF Full-Text] **JNL**

---

283 **Image filtering techniques and VLSI architectures for efficient data extraction in shell rendering**
Olstad, B.; Steen, E.; Halaas, A.
Image Processing, 1995. Proceedings., International Conference on , Volume: 2 , 199
Page(s): 113 -116 vol.2

[Abstract] [PDF Full-Text] **CNF**

---

284 **A programmable motion estimator for a class of hierarchical algorithms**
Lin, H.-D.; Anesko, A.; Petryna, B.; Pavlovic, G.
VLSI Signal Processing, VIII, 1995. IEEE Signal Processing Society [Workshop on] , 1995
Page(s): 411 -420

[Abstract] [PDF Full-Text] **CNF**

---

285 **Prairie: A rule specification framework for query optimizers**
Das, D.; Batory, D.
Data Engineering, 1995. Proceedings of the Eleventh International Conference on , 1995
Page(s): 201 -210

[Abstract] [PDF Full-Text] **CNF**

---

286 **Diversity receiver of a CDMA system in mobile multipath fading environments**
Yang, J.; Sood, P.
Vehicular Technology Conference, 1995 IEEE 45th , Volume: 2 , 1995
Page(s): 684 -688 vol.2

[Abstract] [PDF Full-Text] **CNF**

---

287 **Finding what you want: new tools and tricks**
Dumais, S.
IEEE Software , Volume: 12 Issue: 5 , Sept. 1995
Page(s): 79 -81, 86

[Abstract] [PDF Full-Text] **JNL**

288 **Text retrieval-a trendy cocktail to address the dataworld**
Chellappa, M.; Kambhampaty, S.
Computer Software and Applications Conference, 1994. COMPSAC 94. Proceedings., Eighteenth Annual International , 1994
Page(s): 276 -280

[Abstract] [PDF Full-Text] **CNF**

289 **Ultimedia Manager: Query By Image Content and its applications**
Barber, R.; Flickner, M.; Hafner, J.; Lee, D.; Niblack, W.; Petkovic, D.; Ashley, J.; McConnell, T.; Ho, J.; Jang, J.; Berkowitz, D.; Yanker, P.; Vo, M.; Haas, D.; Lassig, D. Tate, S.; Chang, A.; van Houten, P.; Chang, J.; Petersen, T.; Lutrell, D.; Sned
Compcon Spring '94, Digest of Papers. , 1994
Page(s): 424 -429

[Abstract] [PDF Full-Text] **CNF**

290 **Scalable combinatorial search engine**
Wang, C.-C.; Hamieson, L.H.
Computer Architectures for Machine Perception, 1993. Proceedings , Nov. 1993
Page(s): 342 -351

[Abstract] [PDF Full-Text] **CNF**

291 **Associative processing with CAMs**
Stas, S.
Northcon/93. Conference Record , 1993
Page(s): 161 -167

[Abstract] [PDF Full-Text] **CNF**

292 **The Volcano optimizer generator: extensibility and efficient search**
Graefe, G.; McKenna, W.J.
Data Engineering, 1993. Proceedings. Ninth International Conference on , 1993
Page(s): 209 -218

[Abstract] [PDF Full-Text] **CNF**

293 **Implementation and evaluation of a parallel text searcher for very large text databases**
Hollaar, L.A.
System Sciences, 1992. Proceedings of the Twenty-Fifth Hawaii International Conference on , Volume: i , 1991
Page(s): 300 -307 vol.1

[Abstract] [PDF Full-Text] **CNF**

294 **Massively parallel information retrieval for wide area information servers**
Stanfill, C.

Systems, Man, and Cybernetics, 1991. 'Decision Aiding for Complex Systems, Conference Proceedings., 1991 IEEE International Conference on , 1991
Page(s): 679 -682 vol.1

[Abstract] [PDF Full-Text] **CNF**

---

295 **NRT (News Retrieval Tool): a user's guide**
Sanderson, M.
Large Databases in Press and Publishing: The Present and the Future, IEE Colloquium on , 1990
Page(s): 5/1 -514

[Abstract] [PDF Full-Text] **CNF**

---

296 **A range searching engine for VLSI design CAE**
Chang, S.-S.; Rose, K.
VLSI Technology, Systems and Applications, 1989. Proceedings of Technical Papers. 1989 International Symposium on , 1989
Page(s): 291 -296

[Abstract] [PDF Full-Text] **CNF**

---

297 **A ternary content addressable search engine**
Wade, J.P.; Sodini, C.G.
Solid-State Circuits, IEEE Journal of , Volume: 24 Issue: 4 , Aug. 1989
Page(s): 1003 -1013

[Abstract] [PDF Full-Text] **JNL**

---

298 **Real-time string search engine LSI for 800-Mbit/sec LANs**
Yamada, H.; Murata, Y.; Maeda, T.; Ikeda, R.; Motohashi, K.; Takahashi, K.
Custom Integrated Circuits Conference, 1988., Proceedings of the IEEE 1988 , 1988
Page(s): 21.6/1 -21.6/4

[Abstract] [PDF Full-Text] **CNF**

[Prev] 1 2 3 4 5 6 7 8 9 10 11 12

Home | Log-out | Journals | Conference Proceedings | Standards
Search by Author | Basic Search | Advanced Search | Join IEEE | Establish a Web Account

Your search matched **298** of **676793** documents.
Results are shown **25** to a page, sorted by **publication year** in **descending** order.
You may refine your search by editing the current search expression or entering a new one the text box.
Then click **Search Again**.

search engines

Search Again

Results:
Journal or Magazine = **JNL** Conference = **CNF** Standard = **STD**

226 **A distributed digital library architecture incorporating different index styles**
McNab, R.J.; Witten, I.H.; Boddie, S.J.
Research and Technology Advances in Digital Libraries, 1998. ADL 98. Proceedings. IEEE International Forum on , 1998
Page(s): 36 -45

[Abstract] [PDF Full-Text] **CNF**

227 **Web warehousing: an algebra for web information**
Ng, W.-K.; Lim, E.-P.; Huang, C.-T.; Bhowmick, S.; Qin, F.-Q.
Research and Technology Advances in Digital Libraries, 1998. ADL 98. Proceedings. IEEE International Forum on , 1998
Page(s): 228 -237

[Abstract] [PDF Full-Text] **CNF**

228 **The integration of the World Wide Web and intranet data resources**
Noah, W.W.
System Sciences, 1998., Proceedings of the Thirty-First Hawaii International Conferen on , Volume: 4 , 1998
Page(s): 496 -502 vol.4

[Abstract] [PDF Full-Text] **CNF**

229 **Intelligent Internet searching engine based on hybrid simulated annealin**
Yang, C.C.; Yen, J.; Chen, H.
System Sciences, 1998., Proceedings of the Thirty-First Hawaii International Conferen on , Volume: 4 , 1998
Page(s): 415 -422 vol.4

[Abstract] [PDF Full-Text] **CNF**

230 **Using metadata to query passive data sources**
Martin, P.; Powley, W.; Weston, A.
System Sciences, 1998., Proceedings of the Thirty-First Hawaii International Conferen on , Volume: 7 , 1998
Page(s): 286 -294 vol.7

[Abstract] [PDF Full-Text] **CNF**

231 **Mondou: interface with text data mining for Web search engine**
Kawano, H.; Hasegawa, O.
System Sciences, 1998., Proceedings of the Thirty-First Hawaii International Conferen on , Volume: 5 , 1998
Page(s): 275 -283 vol.5

[Abstract] [PDF Full-Text] **CNF**

232 **A default extension to description logics for use in an intelligent search engine**
Lambrix, P.; Shahmehri, N.; Wahllof, N.
System Sciences, 1998., Proceedings of the Thirty-First Hawaii International Conferen on , Volume: 5 , 1998
Page(s): 28 -35 vol.5

[Abstract] [PDF Full-Text] **CNF**

233 **Comparing genetic algorithms and greedy heuristics for adaptation problems**
Bilchev, G.; Olafsson, H.S.
Evolutionary Computation Proceedings, 1998. IEEE World Congress on Computational Intelligence., The 1998 IEEE International Conference on , 1998
Page(s): 458 -463

[Abstract] [PDF Full-Text] **CNF**

234 **A 1.5-W single-chip MPEG-2 MP@ML video encoder with low power motio estimation and clocking**
Mizuno, M.; Ooi, Y.; Hayashi, N.; Goto, J.; Hozumi, M.; Furuta, K.; Shibayama, A.; Nakazawa, Y.; Ohnishi, O.; Shu-Yu Zhu; Yokoyama, Y.; Katayama, Y.; Takano, H.; Miki, N.; Senda, Y.; Tamitani, I.; Yamashina, M.
Solid-State Circuits, IEEE Journal of , Volume: 32 Issue: 11 , Nov. 1997
Page(s): 1807 -1816

[Abstract] [PDF Full-Text] **JNL**

235 **From bookmark managers to distributed indexing: an evolutionary way to the next generation of search engines**
Kosmynin, A.
IEEE Communications Magazine , Volume: 35 Issue: 6 , June 1997
Page(s): 146 -151

[Abstract] [PDF Full-Text] **JNL**

236 **A simple algorithm for nearest neighbor search in high dimensions**
Nene, S.A.; Nayar, S.K.
Pattern Analysis and Machine Intelligence, IEEE Transactions on , Volume: 19 Issue: 9 Sept. 1997
Page(s): 989 -1003

[Abstract] [PDF Full-Text] **JNL**

237 **Lycos: design choices in an Internet search service**
Mauldin, M.I.
IEEE Expert [see also IEEE Intelligent Systems] , Volume: 12 Issue: 1 , Jan.-Feb. 199
Page(s): 8 -11

[Abstract] [PDF Full-Text] **JNL**

238 **Visually searching the Web for content**
Smith, J.R.; Shih-Fu Chang
IEEE Multimedia , Volume: 4 Issue: 3 , July-Sept. 1997
Page(s): 12 -20

[Abstract] [PDF Full-Text] **JNL**

239 **A tool for organizing Web information**
Mahalingam, K.; Huhns, M.N.
Computer , Volume: 30 Issue: 6 , June 1997
Page(s): 80 -83

[Abstract] [PDF Full-Text] **JNL**

240 **Comparing Internet search engines**
Kingoff, A.
Computer , Volume: 30 Issue: 4 , April 1997
Page(s): 117 -118

[Abstract] [PDF Full-Text] **JNL**

241 **Using information density to navigate the Web**
Ellman, J.; Tait, J.
Intelligent World Wide Web Agents (Digest No.: 1997/118), IEE Colloquium on , 1997
Page(s): 8/1 -8/6

[Abstract] [PDF Full-Text] **CNF**

242 **A component search engine model on Internet**
Ruan Tong; Sheng Haolin
Technology of Object-Oriented Languages, 1997. TOOLS 24. Proceedings , 1998
Page(s): 393 -396

[Abstract] [PDF Full-Text] **CNF**

243 **Intelligent agents for matching information providers and consumers on the World-Wide-Web**
Lee, J.K.W.; Cheung, D.W.; Kao, B.; Law, J.; Lee, T.
System Sciences, 1997, Proceedings of the Thirtieth Hwaii International Conference o
Volume: 4 , 1997
Page(s): 189 -199 vol.4

[Abstract] [PDF Full-Text] **CNF**

244 **HyperSQL: web-based query interfaces for biological databases**
Newsome, M.; Pancake, C.; Hanus, J.
System Sciences, 1997, Proceedings of the Thirtieth Hwaii International Conference o
Volume: 4 , 1997
Page(s): 329 -339 vol.4

[Abstract] [PDF Full-Text] **CNF**

245 **Virtual proxy servers for WWW and intelligent agents on the Internet**
Sun Wu; Chang-Chain Liao
System Sciences, 1997, Proceedings of the Thirtieth Hwaii International Conference o
Volume: 4 , 1997
Page(s): 200 -209 vol.4

[Abstract] [PDF Full-Text] **CNF**

246 **Intelligent spider for Internet searching**
Hsinchum Chen; Yi-Ming Chung; Ramsey, M.; Yang, C.C.; Pai-Chun Ma; Yen, J.
System Sciences, 1997, Proceedings of the Thirtieth Hwaii International Conference o
Volume: 4 , 1997
Page(s): 178 -188 vol.4

[Abstract] [PDF Full-Text] **CNF**

247 **Do images mean anything?**
Santini, S.; Jain, R.
Image Processing, 1997. Proceedings., International Conference on , Volume: 1 , 199
Page(s): 564 -567 vol.1

[Abstract] [PDF Full-Text] **CNF**

248 **Exploring image functionalities in WWW applications development of image/video search and editing engines**
Shih-Fu Chang; Smith, J.R.; Meng, H.J.
Image Processing, 1997. Proceedings., International Conference on , Volume: 3 , 199
Page(s): 1 -4 vol.3

[Abstract] [PDF Full-Text] **CNF**

249 **Integrated Search Engine**
Adah, S.; Bufi, C.; Temtanapat, Y.
Knowledge and Data Engineering Exchange Workshop, 1997. Proceedings , 1997
Page(s): 140 -147

[Abstract] [PDF Full-Text] **CNF**

250 **Experiments in using agent-based retrieval from distributed and heterogeneous databases**

Das, B.; [redacted]ocur, D.
Knowledge and Data Engineering Exchange Workshop, 1997. Proceedings , 1997
Page(s): 27 -35

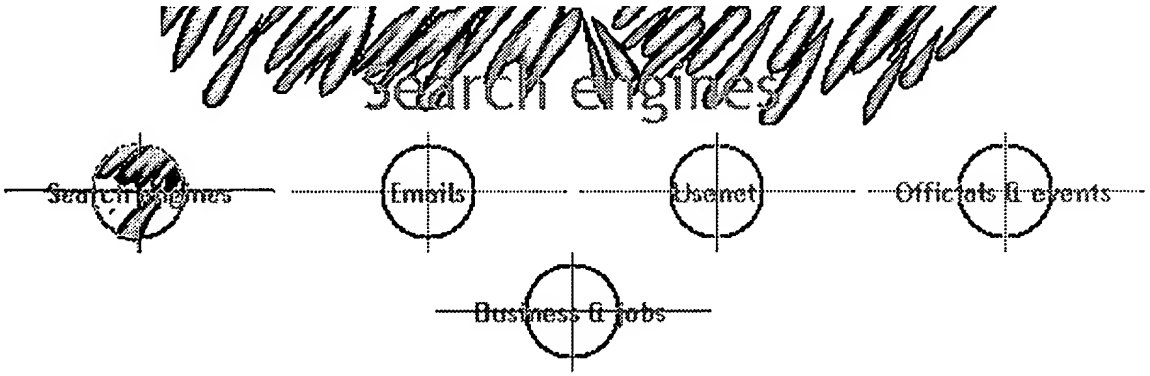
[Abstract] [PDF Full-Text] **CNF**

[Prev] 1 2 3 4 5 6 7 8 9 10 11 12 [Next]

Home | Log-out | Journals | Conference Proceedings | Standards
Search by Author | Basic Search | Advanced Search | Join IEEE | Establish a Web Account

Back to
Teleactivities

Home Up Feedback Contents Search E-mail



Search engines

The level of details in a given story depends in part on the news providers' readers, and the nature of the source. The amount of "noise" (the level of irrelevancy) also varies. In most public forums, expect to wade through many uninteresting messages before finding things of interest.

Try the following strategy:

Step 1: Locate sources that provide relevant information,

Selecting sources is half the battle in making a good search!
You probably won't find what you need if you're not looking in the right place.

Step 2: Check if the information from these sources is at a satisfactory level of details, and that the volume is acceptable (not too much, nor too little).

Step 3: Study the service's search commands and procedures, PLAN, and then SEARCH.

Locating interesting sources

Step 1 is not an easy one. There is such an abundance of directory services and pointers.

On the Internet, two free favorite starting points are Digital Equipment Corp.'s Alta Vista service, and HotBot.

The **Alta Vista** search service indexes millions of Web pages, and maintains a full-text index of

more than 8,000 Usenet newsgroups updated in real- time. Its Advanced Option lets you limit a search by giving start and end dates, by combining words and phrases using AND, OR, NOT, and NEAR operators.

Alta Vista also lets you use a plus sign (+) to include words or a minus sign (-) to exclude words in the search, as in *+online +world -computer*. This search will only return hits containing the words "online" and "world" but not "computer".

Check out Alta Vista at http://www.altavista.digital.com/ (USA), or some of its **mirrors** (local copies of the service) around the world for speed. It offers searches in over 25 languages.

> *It's only worth using Alta Vista if you bear in mind the sort of material which might be posted in your subject area. Since anyone can publish almost anything on the Web, pages vary - from personal pages set up by any student who has Internet access, to those set up academic or research institutions, those set up by not-for-profit organizations, and those from commercial organizations.*

In early 1998, **HotBot** (http://www.hotbot.com) claimed an index of 110 million full-text Web pages, plus Usenet newsgroups and selected Internet mailing lists. This is far more than Alta Vista has, and in some cases it will let you find more.

HotBot supports Boolean AND/OR/NOT, and phrase searching. It provides relevance feedback with retrieval. It also supports chronological, domain, and geographic searches, as well as media type searches such as Java, VRML, and Acrobat, but does not have as powerful search features as Alta Vista.

Watch these strong competitors:

http://www.excite.com
http://infoseek.go.com/.

Meta-searching

Meta-search agents let you search several search engines in one operation. For example, Super Searches (http://www.searches.com/) searches major search engines like Alta Vista, Excite, Galaxy, HotBot, Lycos, Web Crawler, Yahoo, WWW Yellow pages, Meta crawler, Deja.com, Aliweb, Hotbot, Lycos, and more.

Here are some others to try:

Dogpile: http://www.dogpile.com

Highway61: http://www.highway61.com/

One word of warning: The meta-search agents treat the product of search engines as data: changing it, organizing it, and making it simpler to use for the consumer, without understanding that this information is more like a publication than raw data.

Usually, these services do not support Boolean, temporal, or proximity operators. Set building is not possible.

Searching a topic area

Narrowing a search down to a specific topic area can be a challenge with the general search engines. Sometimes, you may be better off using a more targeted search service.

There are many services linking you to topic area search engines. Example: **SEARCH.COM** (http://www.search.com/) links you to search services within areas like Arts, Automotive, Business, Computers, Directories, Education, Employment, Entertainment, Finance, Government, Games, Health, Housing, Legal, Lifestyle, News, People, Politics, Reference, Science, Shopping, Sports, Travel, Usenet, and Web.

Langenberg Search (http://www.langenberg.com/) is a gateway to some of the most popular search engines for a variety of subjects grouped under : Acronym, Area Codes, Books&Pubs, BusinessFinder, Cooking, Dictionary, Encyclopedia, Entertainment, Government, Jobs, Maps, Medicine, Metasearch, Misc, Money&Stocks, News&Sports, PersonFinder, Religion, SearchEngines, Shipping, Translation, Travel, Usenet, Weather, Zip Codes. **The BIG Search Engine Index** (http://www.merrydew.demon.co.uk/search.htm) may also be worth your visit.

Some other interesting offerings:

http://www.newsindex.com/	Today's news.
http://www.newstrawler.com	Archives of yesterday's news
http://www.uni-karlsruhe.de/~un9v/atm/ase.html	Airport Search Engine
http://www.cyberark.com/noah.htm	Animals
http://www.ciolek.com/SearchEngines.html#asia	Asia (Search Asian Studies WWW VL Web Space)
http://www2.zdnet.com/locator/, and http://www.computercrowsnest.com/	Computer companies, hardware, software, peripherals.
http://www.webplaces.com/search/	Clip art, icons, background images, animations, sound clips
http://www.search.com/Single/0,7,0-300506,0200.html	Education
http://www.financewise.com/	Financial-only content
http://www.faqs.org/faqs/faqsearch.html	Frequently Asked Questions (FAQs)
http://thegw.com and http://www.pcgame.com	Games on the Net.
http://www.achoo.com	Health
http://www.Healthatoz.com	Health
http://www.sprocketsandcogs.com/	Html, dhtml, Perl, Java: for Web developers and programmers.

http://www.topica.com	Mailing lists by topic area
http://www.scour.net http://www.lycos.com/picturethis/ http://image.altavista.com/cgi-bin/avncgi http://www.arribavista.com/	Multimedia files such as movie scenes, pictures, music clips, concerts, sporting events.
http://mp3.lycos.com/	Music (MP3 format)
http://www.searchz.com	Marketing: For Online Advertisers, Marketers, and E-Commerce
http://www.idealist.org/	Directory & Search Engine For Non-Profit Organizations
http://www.isinet.com/	Scientific information.
http://www.simtel.net/simtel.net/ http://ftpsearch.lycos.com/ http://www.tucows.com/	Software - shareware and public domain
http://www.nlsearch.com/	Web searches. In addition, searches in million of articles from 5,400 premium sources, such as books, magazines, databases, and newswires not available elsewhere.

Searching for non-US information

No search engine indexes the whole Web, and most US based services tend to be best at US contents. US services focusing on other geographical areas tend to miss local organizations having registered .com, .org, or other global addresses.

For contents in other geographical areas, you may be better served by engines specialized on these areas. Examples:

Europe	http://www.euroferret.com/ http://www.euroseek.net/
India	http://www.geocities.com/SouthBeach/4195/india.htm
India/Pakistan/ Sri Lanka/Nepal/ Bangladesh	http://www.samilan.com/
Israel	http://www.vci.co.il/
Middle East	http://www.arabseek.net/
Russia	http://search.interrussia.com/
Scandinavia	http://www.polarsearch.com/
South Africa	http://www.ananzi.co.za/
United Kingdom	http://www.mirago.co.uk/

For links in other countries, try **Search Engines Worldwide** at http://www.twics.com/~takakuwa/search/, and http://www.beaucoup.com/1geoeng.html.

Non-English language searches

There are major structural differences between languages. An indexing system built for English text may therefore not be suitable for a text written in the language you're searching, and in particular if the other language uses special fonts. Using special purpose search engines may be the way to go in such cases. Some options:

Arabic	http://www.alidrisi.com/main1.htm
Chinese	http://www.sohoo.com.cn/Computer/Internet/Search/index.html
French	http://lokace.iplus.fr/ http://www.ecila.fr
German	http://www.aladin.de/ http://www.dino-online.de/suche.html
Italian	http://ragno.ats.it/indexuk.html
Japanese	http://www.lawresearch.com/v2/Cejapan.htm
Spanish	http://www.ctv.es/USERS/gobib/hispano.html

Another problem using the English language search systems is that you don't just have to understand English to get the most out of them, you'll have to understand English well.

More sources about sources

Scott Yanoff updates an interesting, selected list of Internet resources twice per month. Get it by email from inetlist@aug3.augsburg.edu, or from

http://www.spectracom.com/islist/
ftp://ftp.csd.uwm.edu/pub/inet.services.txt

John December's "Information Sources: the Internet and Computer-Mediated Communication" has pointers to information describing the Internet, computer networks, and issues related to computer-mediated communication. It lists Internet texts for new users, comprehensive Internet guides, and specialized and technical information. At http://www.december.com/cmc/info/index.html

The Gale Directory of Databases contains detailed descriptions of over 11,500 publicly available databases accessible through an online vendor or batch processor or for purchase on CD-ROM, diskette, or magnetic tape, or as a handheld product (Feb, 1999). It is a comprehensive guide to the electronic database industry worldwide.

The directory is available in print, on CD-ROM, through Dialog and other commercial services, and through Gale Research's subscription-based Web service (at http://www.gale.com/). They also offer listings of database producers and vendors.

For lists of electronic journals about the Internet ("E-zines" or "Ejournals"), click at http://www.edoc.com/ejournal/

Several electronic journals and newsletters are available through the Internet, covering fields from literature to molecular biology. For a large list, try http://www.meer.net/~johnl/e-zine-list/.

The NEWSLTR list distributes various network newsletters. Subscribe by email to listserv@listserv.nodak.edu. Offerings include: Edupage, Hitek, HPC, Infosys, IAT Inforbit, and many more.

The Argus Clearinghouse offers over 1,000 topical guides to the Internet's information resources. The guides are created by librarians and other information professionals, and cover a diverse range of topics, from Theatre, Law, and Chemistry to Midwifery. Access on this Web address: http://www.clearinghouse.net/

Interested in CD-ROM? The database at http://www.microinfo.co.uk/ offers details about thousands of information products and services - mainly CD- ROMs. Products are classified in 27 topics ranging from agriculture and food to theology.

Practical hints about online searching

We cannot give a simple, universal recipe valid for all online services. The best approach on one service, may be useless on others.

Besides, recommendations will vary considerably depending on whether you want "focused searches" designed to find and retrieve a specific set of documents providing a specific set of information, or "satisfied searches" designed to find just some hits that are "good enough" regardless of the source.

On some services, searching starts by selecting databases or type of source. This may help you get rid of some irrelevancies. On other services, this selection is assumed.

The next step is to enter your search words (or text strings), and a valid time frame (as in "between 1/1/90 and 1/1/91"), where such an option is available.

Here are some sample search terms used on the net:

SONY AND VIDEO	The term SONY **and** the term VIDEO. Both words must be present in the document to give a match.
VIDEO*	search for all words starting with VIDEO. "*" is a wild-card character referring to any ending of the word. VIDEO* matches words like VIDEOTEXT and VIDEOCONFERENCE.
SONY WITHIN/10 VIDEO	Both words must be present in the text, but they must not be farther apart than ten words. (Proximity operators)
IBM OR APPLE	Either one word OR the other.

Some services have adjacency operators, and some automatic truncation. Truncation allows searching on different word endings or plurals with the use of a truncation wild card symbol. For example, if the truncation symbol is *, then the search term econ* will return items that contain economics, economy, economic, and econometric. Car* will return items that contain cars and cartoon, so it is advisable to use truncation symbols carefully.

Many services let you reuse your search terms in new search commands. This may save you time (and money), when you get too many hits. For example: if IBM OR APPLE gives 1,000 hits, limit the search by adding "FROM JANUARY 1st.," or by adding the search word "NOTEBOOK*".

Most services offer full online documentation of their search commands. You can read the help text on screen while connected, or retrieve it for later study. Expect the quality of these texts to be variable, but browse them all the same.

Make a note about the following general tricks:

The use of ANDs and ORs

is called Boolean searching. It allows search terms to be put into logical groups by the use of connective terms.

Using **AND**, **OR**, and **NOT** search operators may seem confusing at first, unless you already understand the logic. Here are some hints that you may find helpful:

Use the Boolean operator AND to retrieve smaller amounts of information. Use AND when multiple words must be present in your search results (MERCEDES AND VOLVO AND CITROEN AND PRICES).

Use OR to express related concepts or synonyms for your search term (FRUIT OR APPLES OR PEARS OR BANANAS OR PEACHES).

The purpose of NOT is avoid listings of irrelevant records. Be careful when using this operator.

NOT gets rid of any record in a database that contains the word that you've "noted" out. For example, searching for "IBM NOT APPLE" drops records containing the sentence, "IBM and Apple are computer giants." The record will be dropped, even if this is the only mention of Apple in an article, and though it is solely about IBM.

Use NOT to drop sets of hits that you have already seen. Use NOT to exclude records with multiple meanings, like "CHIPS Not POTATO" (if you are looking for chips rather than snack foods).

Often, it pays to start with a "quick-and-dirty" search by throwing in words you think will do the trick. Then, look at the first five or 10 records, but look only at the headline and the indexing. This will show you what terms are used by indexers to describe your idea and the potential for confusion with other ideas.

Use proximity operators to search multiword terms. If searching for "market share," you want the two words within so many words of another. The order of the words, however, doesn't matter. You can accept both "market share" and "share of the market."

Relevance ranking, and more

Some claim that boolean searches only find between 20 - 25 percent of the relevant information. The problem is that you must know the terms to search on before you begin. Many people don't know these terms and cannot guess them.

Several online services are busy trying to supply better "search engines" using techniques like natural language searching, relevance ranking, and concept searching.

Relevance ranking tries to measure how closely the retrieval matches the query, usually in quantitative terms between 0 and 100 or 0 and 1,000. It usually provides a ranked listing of search results, with a score for the relevance of the result, based on the occurrences of the terms used and also their position in the document. It provides somewhat the same results as AND searching. Also, it offers the benefits of OR searching as all the terms in a query need not be present in the result.

Alta Vista (http://www.altavista.digital.com/) offers both boolean and enhanced relevance ranking searches. For example, you can require that selected terms be found in the results. The query "+apples +bananas oranges" will not find a document missing the words apples and bananas. Those files that contain oranges will listed before those that do not contain this word, but files without this word will also be listed.

Some services let you search specific types of information. For example, Alta Vista allows searches for characters or words in an URL (a Web address), or a hyperlink.

> Application: My Web pages are at http://home.eunet.no/~presno/. The query "+link:eunet.no/~presno/ -url:eunet.no/~presno/" will most likely find all links to my pages on other Web servers except my own. The "-" character in front of a word works as a NOT operator. The "link:" phrase is for searching in hyperlinks across the Internet. The "url:" code lets you search in the URL addresses of the found pages.

Key Word In Context (KWIC) searching will return the key word and N words near the key word to give the user the context in which the key word was found.

Phrase Searching allows searching of phrases when available. Note that some systems can be confusing if you think "Online World" is searching the two words together as a phrase, when in fact the engine is searching Online OR World.

Fuzzy searching is another interesting concept. This option allows you to search when you don't know the exact spelling of the word. Some systems use the Soundex algorithm invented over 70 years ago to search name files. Names that sound alike should have the same Soundex number. It uses these basic rules:

- Vowels are ignored.
- Consonants that sound alike in a pronounced name have the same "number".
- Successive consonants with the same number are counted as one (Willitt is equal to Wilith).

Note: The information available in English language may be just a small part of that available in a country's national language. When English language sources fail to meet the need at hand, consider the services of a skilled bilingual searcher.

Spelling errors are very common reasons for search failures. Make sure you have that terminology term or person's name right. Also, names are not spelled the same way in all countries, and those who produce texts also make spelling errors. For example, the name of the composer Tchaikowsky is supposedly spelled in 36 different ways on the nets. 'Ciaikovsky' is one of them.

Internet Searching Tip Sheet

Remembering the different commands for all the Internet search engines is difficult, if not impossible! This tip sheet was created as a guide to help you use some of the best search engines on the Internet. This sheet is a summary of the basic commands used by these search engines. Each search engine offers unique searching features not listed here. To fully use a search engine, be sure to read its help screens and print them for future reference.

The table below lists the name of the search engine; the term links/logic operators in use; if phrase searching is available and the symbol used; if searching for plurals is available and the symbol used; if the results are ranked by relevancy; if Helper Boxes are available; and if searches can be restricted by date.

Search Engine	Term Links	Phrases	Plurals	Relevancy Ranking	Helper Boxes	Date Restriction
AltaVista	+ -	" "	*	Automatic	Limited	No
AltaVista Advanced	AND, OR, NEAR, ( )	" "	*	Custom	Limited	Yes
Deja News	AND, OR, NEAR, ( )	" "	*	No	Limited	No
Deja News Power Search	AND, OR, NEAR, ( )	" "	*	Custom	Extensive	Yes
Dogpile	AND, OR, NEAR, ( )	" "	No	Automatic	Limited	No
Excite	AND, OR, ( ), +-, Automatic Phrase Searching	Automatic	Some Automatic	Automatic	No	No
Excite Power Search	Helper Boxes	Option	Some Automatic	Automatic	Extensive	No
HotBot	+ -, Helper Boxes	" "	*	Automatic	Extensive	Limited
HotBot Supersearch	+ -, ( ), AND, OR; Helper Boxes	" "	*	Automatic	Extensive	Yes
Northern Light	Automatic Phrase Searching, + -, AND, OR, ( )	" "	*	Custom	No	No
Northern Light Publication Search	Automatic Phrase Searching, + -, AND, OR, ( )	" "	*	Custom	Limited	No
Northern Light Industry Search	Automatic Phrase Searching, + -, AND, OR, ( )	" "	*	Custom	Extensive	Yes
Northern Light Power Search	Automatic Phrase Searching, + -, AND, OR, ( )	" "	*	Custom	Extensive	Yes
Yahoo	OR, AND, + -	" "	*	Yes	No	No
Yahoo Advanced	+ -	" "	*	Yes	Extensive	Limited

Net Search Tools/Indices/Locators

- All-in-One Search Page: Site features a collection of search engines covering almost any topic, including people, news, world wide web, reference, publications and more.
- Alta Vista: One of the largest and most powerful search engines; particularly useful for complicated searches. Also offers a free translation service, a full-text index of Usenet newsgroup archives and a business and residential telephone directory.
- Ask Jeeves: Ask questions in natural language and identify web sites that might have the answer
- Direct Search: This page compiles links to over 800 specialized and other interactive tools for finding information traditional search engines can't uncover.
- Dogpile: Highly rated meta search engine consisting of 13 WWW Search engines, 6 Usenet sources, and 2 FTP archives.
- EINet Galaxy Directory: EINet Galaxy: subject catalog and search tools.
- Excite: Highly rated search engine featuring Boolean searching, relevancy ranking, and alternative word suggestions. The power search page offers powerful advanced searching by structuring a Boolean search with fill-in boxes.
- FINDSPOT: Tips on how to conduct an effective search using web search utilities (Excite), meta search utilities (MetaCrawler), Usenet search utilities (AltaVista), web directories (Yahoo) and internet resource directories (e-mail addresses).
- Finding People on the Net: This site offers a collection of tools to locate an individual's email address, phone number, or street address. Also offers resources for locating U.S. and European businesses and organizations.
- Gopher Jewels: Gopher jewels--the best gopher sites, sorted by category. Excellent business resource section.
- HotBot: Highly rated search engine for Web or Usenet groups. HotBot ranks results for relevancy, allows complex searches in a simple interface. Allows for specific searches in:: classifieds, domain names, discussion groups, shareware, e-mail addresses, audio recordings and visual images.
- Infoseek: Infoseek (includes Usenet newsgroups and non-internet databases) Offers additional databases for low fee, including wire services.
- Internet Resource Guide Directory: The Argus Clearinghouse is a categorized rated directory of business, technical and personal sites on the Internet.
- Internet Sleuth: The Internet Sleuth (search tool) "A collection of over 900 searchable databases on the internet on a wide variety of subjects."
- Livelink Pinstripe: Internet search engine designed specifically for the business user with slicing technology that allows the user to go directly to a particular business topic.
- Lycos: A well regarded search engine offers a rating system (the top 5% of Web sites) helping to retrieve more valuable information. Also features a $20 Dun & Bradstreet report and the ability to search the Web for sounds and pictures.
- MetaCrawler: One of the best meta search engines, this one allows searching for an exact phrase and offers a single interface for nine search engines, including AltaVista, Infoseek, Lycos and Excite.
- Newest Internet Resources: Highlights the newest internet resources and announcements "verified for substantial content and accessibility..."
- Northern Light: Excellent new low-cost online database vendor serves up high-value publications along with a simultaneous Web search. Business, technical, and general interest periodicals are included at this advanced, yet easy-to-search site.
- Search Engine Showdown: Summaries, reviews, and comparisons of the search features

and database scope of Internet search engines and finding aids.

- Telephone Directories: Telephone Directories on the Web (International)
- Web Site Rankings Directory: Point Communications. Commercial site featuring web site reviews: "Top 5% of All Web Sites" rankings. Contains an excellent list of newspapers on the web.
- Webcrawler: Webcrawler. Document content-based retrieval in addition to title and URL.
- Yahoo!: One of the largest and most popular Internet directories, Yahoo! also sports a useful search engine. An excellent, vast site with multiple uses such as searches for job postings, company listings, news, etc.

Home Up Next

Send mail to sfetcu@teleactivities.net with questions or comments about this web site.

Last modified: December 30, 2000

WEST

Help | Logout | Interrupt

Main Menu | Search Form | Posting Counts | Show S Numbers | Edit S Numbers | Preferences

Search Results -

Terms	Documents
multiple search engines and (kwic or key word context)	1

Database: US Patents Full-Text Database; US Pre-Grant Publication Full-Text Database; JPO Abstracts Database; EPO Abstracts Database; Derwent World Patents Index; IBM Technical Disclosure Bulletins

Refine Search: multiple search engines and (kwic or key word context) Clear

Search History

Today's Date: 4/5/2001

DB Name	Query	Hit Count	Set Name
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	multiple search engines and (kwic or key word context)	1	L4
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	multiple search engines and kwic or key word context	16	L3
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l1 and kwic or key word context	16	L2
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	multiple search engines	22	L1

WEST

Help | Logout | Interrupt

Main Menu | Search Form | Posting Counts | Show S Numbers | Edit S Numbers | Preferences

Search Results -

Terms	Documents
149 and review	55

Database: US Patents Full-Text Database; US Pre-Grant Publication Full-Text Database; JPO Abstracts Database; EPO Abstracts Database; Derwent World Patents Index; IBM Technical Disclosure Bulletins

Refine Search: Clear

Search History

Today's Date: 4/5/2001

DB Name	Query	Hit Count	Set Name
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	149 and review	55	L50
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	148 and list	115	L49
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	12 and movie or film titles	240	L48
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	143 and (movie or film titles)	1	L47
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	143 and movie or film titles	125	L46
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	143 and movies or films	1731477	L45
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	143 and movies	0	L44
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	140 and titles	8	L43
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	140 and movie	1	L42
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	140 and movie title	0	L41
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	139 and list	10	L40
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	138 and terms	11	L39
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	137 and query	11	L38
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	12 and meta information	13	L37
USPT	5864871.pn.	1	L36

USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l2 and jump links	2	L35
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l2 and jump same links	45	L34
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l32 and jump	2	L33
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l28 and links	25	L32
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l28 and skip same links	0	L31
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l28 and jump same links	0	L30
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l28 and highlight	4	L29
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l26 and query same terms	33	L28
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l26 and highlight	11	L27
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l2 and filtering same documents	69	L26
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l2 and clustering documents	7	L25
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l21 and (special characters or letters)	2	L24
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l21 and special characters or letters	270536	L23
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l20 and html tags	31	L22
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l19 and (html comments or annotations)	2	L21
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l19 and html comments or annotations	2975	L20
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l18 and repeated	97	L19
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l17 and characters	141	L18
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l16 and redundant	194	L17
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l2 and whitespace or blank space	4019	L16
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l14 and context same strings	2	L15
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l7 and filtering	92	L14
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l7 and filtering same context same strings	0	L13
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l7 and filtering same pages	11	L12
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l10 and query same terms	15	L11
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l8 and information	50	L10
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l7 and information	405	L9
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l7 and parsing	50	L8
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l2 and third party or multiple party	476	L7
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l5 and query same terms	13	L6
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l4 and information	38	L5
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l3 and parsing	38	L4
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	l2 and third party	137	L3
USPT,PGPB,JPAB,EPAB,DWPI,TDBD	search engines	1392	L2
USPT	search engines	1065	L1

WEST

Generate Collection

L2: Entry 1 of 16 File: USPT Aug 24, 1999

US-PAT-NO: 5943669
DOCUMENT-IDENTIFIER: US 5943669 A

TITLE: Document retrieval device

DATE-ISSUED: August 24, 1999

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Numata; Kenichi	Nakai	N/A	N/A	JPX

ASSIGNEE-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY	TYPE CODE
Fuji Xerox Co., Ltd.	Tokyo	N/A	N/A	JPX	03

APPL-NO: 8/ 975468
DATE FILED: November 21, 1997

FOREIGN-APPL-PRIORITY-DATA:

COUNTRY	APPL-NO	APPL-DATE
JP	8-313761	November 25, 1996
JP	9-166100	June 23, 1997

INT-CL: [6] G06F 17/30
US-CL-ISSUED: 707/5; 707/3, 707/4, 707/500, 707/513, 704/9
US-CL-CURRENT: 707/5; 704/9, 707/3, 707/4, 707/500, 707/513
FIELD-OF-SEARCH: 707/4, 707/5, 707/3, 707/500, 707/513, 704/9

PRIOR-ART-DISCLOSED:

U.S. PATENT DOCUMENTS

Search Selected Search ALL

	PAT-NO	ISSUE-DATE	PATENTEE-NAME	US-CL
☐	5130924	July 1992	Baker et al.	707/5
☐	5325298	June 1994	Gallant	707/5
☐	5434962	July 1995	Kyojima et al.	395/145
☐	5619709	April 1997	Caid et al.	704/9
☐	5724593	March 1998	Hargrave, III et al.	704/7
☐	5794178	August 1998	Caid et al.	704/9

FOREIGN PATENT DOCUMENTS

FOREIGN-PAT-NO	PUBN-DATE	COUNTRY	US-CL
4-84271	March 1992	JPX	

OTHER PUBLICATIONS

Chung-Hsin, Lin et al., "[illegible]automatic indexing and neural ne[illegible]rk approach to concept retrieval and classification of multilingual (Chinese-English) documents", IEEE Transactions on Systems, Man and Cybernetics, Part B, Cybernectics, vol. 26, No. 1, Feb. 1996.
Cunningham, S.J. et al., "Applying machine learning to subject classification and subject description for information retrieval", Proceedings of Second New Zealand International Two-Stream Conference on Artificial Neural Networks and Expert Systems, 199, Nov. 1995.
Legakis, L., "Intelligent subject matter classification and retrieval", Canadian Conference on Electrical and Computer Engineering, 1993., vol. 1, Sep. 1993, pp. 15-18.

ART-UNIT: 276
PRIMARY-EXAMINER: Lintz; Paul R.
ASSISTANT-EXAMINER: Alam; Shahid
ATTY-AGENT-FIRM: Oliff & Berridge, PLC

ABSTRACT:

Parts of documents are retrieved using the entire context of selected documents. Classification unit designation section performs the designation of a classification unit. A logical structure analysis section analyzes the logical structure of the documents read-in from a document storing section where the documents are stored. A fundamental vector generation section partitions the logical structure of the documents by means of the classification unit, extracts keywords, and generates fundamental vectors. A heading vector generation section extracts key words from the headings of the structural elements that are arranged in higher level of structure than the structural element of the classification unit that was the target of fundamental vector generation, and generates heading vectors. A vector synthesis section synthesizes fundamental vectors and heading vectors, and generates composite vectors. Composite vector maintenance section attaches the corresponding composite vectors to structural elements of the classification unit that were the target of composite vector generation and maintains the attached objects. A classification section classifies the structural elements of the documents of the classification unit based on the degree of similarity of the generated composite vectors. A display section displays the results of classification.

16 Claims, 31 Drawing figures

Generate Collection

L6: Entry 3 of 13 File: USPT Nov 21, 2000

DOCUMENT-IDENTIFIER: US 6151624 A
TITLE: Navigating network resources based on metadata

ABPL:
Mechanisms for associating metadata with network resources, and for locating the network resources in a language-independent manner, are disclosed. Owners of network resources define metadata that describes each network resource. The metadata may include a natural language name of the network resource, its location, its language, its region or intended audience, and other descriptive information. The owners register the metadata in a registry. A copy of the metadata is stored on a server associated with a group of the network resources. A copy of the metadata is stored in a registry that is indexed at a central location. A crawler service periodically updates the registry by polling the information on each server associated with registered metadata. To locate a selected network resource, a client provides the name of the network resource to a resolver process. The resolver process provides to the client the network resource location corresponding to the network resource name. Multiple metadata mappings can be established for the same network resource, in which each mapping stores a name expressed in a different natural language. Accordingly, network resources can be located merely by providing the name of the network resource in any natural language that is convenient for the client.

BSPR:
Recently, a global packet-switched network known as the Internet has attracted wide use. A local computer can connect to a distant server, request a file or an image from the server, and receive the requested information immediately.

BSPR:
Accordingly, in 1984 the Domain Name System (DNS) was introduced. DNS is a distributed information database that maps the IP address of a server to a host name or "domain name". For example, the domain name www.centraal.com is mapped to the IP address 209.76.153.3 in the DNS system. The database is available at several computer systems around the world known as DNS servers. A local computer can look up a remote server by connecting to a DNS server, providing a domain name to the DNS server, and obtaining the IP address that corresponds to the domain name. The local computer can then connect to the remote computer using the IP address, and send and receive information.

BSPR:
One popular technology enjoying wide use with the Internet is known as the World Wide Web. The World Wide Web enables a computer to locate a remote server using the DNS and then establish a connection to the server and retrieve information using a communication protocol called the Hypertext Transfer Protocol (HTTP). Each item of information available using the Web, including files, images, or pages, is called a resource. A Uniform Resource Locator (URL) uniquely identifies each resource stored on a server. A URL is a form of network address comprising a domain name coupled to an identifier of the location of information stored in a network.

BSPR:
An example of a URL is http://www.centraal.com/index.html. In this example, "http://" indicates that the information associated with the URL can be accessed using HTTP; www.centraal.com identifies the server that is storing the information; and "index.html" identifies a file or page on that server.

BSPR:
The local computer requests information by providing a request containing a URL of the desired information to the remote server. The server receives the request, locates the page of information corresponding to the URL, and returns the page to the local computer over the HTTP connection. The pages of information are files prepared in the Hypertext Markup Language (HTML). The local computer runs a browser program that can read HTML files, interpret HTML codes in the files, and generate a complex graphical display.

BSPR:

- Because the Web offers so [illegible]ch information about so many sub[illegible]ts, often the Web is compared to a library. In this analogy, the books in the library are network resources such as Web pages. All of the books are written in the same language, namely HTML. Unfortunately, although HTML is a simple language, it does not provide a mechanism that can be used to express attributes relating to a network resource. Thus, continuing the library analogy, a Web page is like a book that has no cover. The content of the Web page can be read, but there is no descriptive information about the Web page, such as its title, subject, or publication date, associated with the Web page. It is difficult to identify or refer to a book that has no title. Since Web pages do not inherently contain a cover that stores a title, conventionally, Web pages are referenced by a location identifier or URL in the DNS system. The current DNS system as implemented with the Web has several disadvantages and drawbacks. Although the DNS system ensures that each URL is unique across the Web, URLs are difficult to remember and associate with a particular institution, person, or product related to the owner of the domain or page associated with the URL. For example, to locate a page of information about the Walt Disney film "Bambi", in the current system a user must enter a complex URL into the browser, such as http://www.disney.com/DisneyVideos/masterpiece/shelves/bambi.

BSPR:
Thus, an inherent disadvantage of the DNS system is that the user must know the exact location and name of the desired information. In the library analogy, URLs are like card catalog numbers. Few persons go to a library knowing the exact card catalog number of a desired book. However, in the Web environment, there is no alternative, even though users tend to naturally remember the names of network resources but not their locations. Moreover, network resources are volatile; their locations may change or be reorganized over time at the discretion of the operator of the server that stores the network resource. Thus, a URL that is accurate one day might be inaccurate the next day, so that the network resource cannot be located.

BSPR:
Because of the difficulty of associating a location identifier with a desired network resource, specialized Web sites known as " search engines" have been developed to provide a way to enter natural language words or phrases and retrieve a list of other Web sites that contain the words or phrases. Examples of search engines are AltaVista, Yahoo!, and Lycos. However, search engine technology has limitations and drawbacks. For example, search engines do not understand the content of the Web pages indexed by the search engine; search engines merely remember the Web pages.

BSPR:
Further, search engines merely return a list of Web pages that contain the words or phrases entered by the user; they do not automatically navigate to a pertinent page. The list returned by the search engine may have thousands of entries, many of which are irrelevant to what the user wants. In the library analogy, this process is like requesting a librarian to search for a book, and receiving from the librarian a list of card catalog numbers at which the book might be located.

BSPR:
In addition, the list almost always contains entries that merely mention the words or phrases entered by the user but are not associated with the owner of a product or service identified by those words or phrases. For example, a user might want to locate the Web site owned and operated by United Airlines. The user enters "United Airlines" into the query field of a search engine. The search engine returns a list of Web sites or Web pages that contain the words "United Airlines." However, many of the entries in the list are not owned or operated by United Airlines; they are owned or operated by third parties that merely mention the words in their pages. Further, the lists produced by search engines often are unordered, so that the user must carefully search the list to identify a desired entry. While search engine technology may have been adequate when the Web contained only a few documents, the Web is currently estimated to contain more than 200 million pages, rendering impractical the continued use of search engines based on location identifiers. Some have proposed making search engines smarter, using new ranking algorithms, semantic analysis, and HTML filtering techniques. Nevertheless, search engine performance continues to degrade because the Web is growing faster than search engine technology is improving.

BSPR:
Search engines also suffer from the disadvantage that they can be fooled by metatags. The HTML language defines a metatag facility whereby text such as key words or descriptions is written into a Web page's HTML code as a means for a search engine to categorize the content of the Web page. The browser does not display the metatags when the Web page is received and decoded at the client. The metatag facility can be used to fool a search engine by encoding a non-displayed keyword into a Web page that has nothing to do with the actual content of the page. When the keyword is used for a Web search, the Web page is located and displayed even though the displayed content of the page is unrelated to the key word.

BSPR:
It is also desirable to have a way to access information available over the Web using a natural language word or "real" name associated with the information.

BSPR:
It is also desirable to have a Web browser program that can rapidly locate, load, and display information in response to receiving a natural language word or "real" name associated with the information, thereby providing a way to instantly retrieve information stored in a network based upon the real name rather than the address of the information.

BSPR:
It is also desirable to have such a system that can automatically and immediately navigate or direct the user to a particular network resource, without providing or requiring the user to search through a list of results or matches. It is also desirable to have a flexible, simple way to associate a natural language word or "real" name with a set of information.

BSPR:
It is also desirable to have a way to associate information stored in a network with human-readable resource names, so that end users can navigate the network using simple words and sentences expressed in any human written language.

BSPR:
It is also desirable to have such a system configured in a way that provides distributed storage of the real name information.

BSPR:
Yet another feature involves the steps of retrieving the name file; parsing the name file; building an index entry based on the values parsed from the name file; and storing the index entry in an index that is stored apart from the storage device. Still another feature is the steps of sending the name file over the network to a client associated with the resource; and storing the name file in a server storage device of a server associated with the client. Another feature involves periodically polling the name file on the server associated with the client; testing whether one of the natural language names stored in the name file matches a third natural language name stored in a database indexed by the index; and updating the database when changes are detected in the name file. Yet another feature is the step of synchronizing the index to the database.

DEPR:
In the preferred embodiment, metadata is associated with network resources such as Web pages. Generally, metadata is data that describes other data. The metadata defined herein provides information that describes a Web page in a manner analogous to the manner by which a catalog card describes a book in a library. For example, the metadata includes information that provides a title (also called a real name address), a description, a language designation, or a geographical location. The metadata is defined by an administrator of the server that stores the Web pages that are described in the metadata, and a copy of the metadata is stored in association with that server so that the metadata is accessible using the Web. Using a Librarian, the a copy of the metadata is registered with a database that is coupled to an index.

DEPR:
Preferably, the metadata is prepared and initially stored in the form of a Name File 64 is a text file defined by the Extensible Markup Language (XML) grammar. XML is a language definition promoted by Microsoft Corporation and Netscape Communications Corporation. Further information about XML is provided in "XML: Principles, Tools, and Techniques," The World Wide Web Journal, vol. 2, no. 4 (Fall 1997) (Sebastopol, Calif.: O'Reilly & Assoc., Inc.).

DEPR:
The RNS file 900 is defined according to a grammar in which information elements are surrounded by complementary tags. For example, "" and "" are complementary tags. The RNS file 900 has two general parts, namely a schema section 902, and a data section 904. The schema section 902 and the data section 904 are enclosed within complementary tags (", ") that indicate that the RNS file 900 is in the XML grammar.

DEPR:
For example, one or more Name Files 64 have entries that store real names in English, French, German, and Japanese. Each entry identifies the same network resource. Accordingly, the entries establish real names in a plurality of different languages, all of which point to or resolve to the same network address. When a third party wishes to access the referenced network resource, the third party enters the real name of the network resource into the browser 74 or the GO service 42 in whatever language is most

convenient for the third party. The Resolver 40 will resolve the real name, regardless of language, to the same network address and direct the browser to that address. Accordingly, a user can locate and access network resources in a language-independent manner.

DEPR:
In an alternative embodiment, the resources described in the Name File 64 are persons rather than Web pages. A resource of type "person" has metadata including a mailing address, email address, and other personal information. In this embodiment, the system can be used as a person locator service rather than for navigating to Web pages or other network resources.

DEPR:
In other alternative embodiments, the Name File 64 stores other attributes. For example, other attributes include Organization, Subject, Abstract, Type, Audience, and other attributes. In the Organization attribute the Name File 64 information that identifies an organization or company that owns or is associated with the network resource, for example, "Federated Stores Incorporated." In the Subject attribute the Name File 64 stores information that describes the subject matter of the network resource, for example, "dogs." In the Abstract attribute the Name File 64 stores information containing an abstract of the network resource. In the Type attribute the Name File 64 stores information describing a type of the network resource, for example, " RealAudio file". In the Audience attribute the Name File 64 stores information describing the intended audience of the network resource, for example, "Women age 19-34".

DEPR:
The Registry 10 includes a database 12 in the form of a commercial database system, such as the SQL Server, or a proprietary database. The Registry 10 provides a centralized storage point for mappings of real names to network addresses or URLs, as well as descriptive information associated with the real names. In this context, "real name" refers to a name of a network resource expressed in conventional syntax of a natural language, such as English, Japanese, Russian, etc. Each real name is required to be unique across the Internet and unique within the Registry 10. The uniqueness of real names is enforced by the Registry 10. The Registry 10 operates as a centralized, highly robust, and scalable persistent storage area for all metadata. The Registry 10 also stores statistics related to the usage of the metadata in the context of various services that are built on top of the Registry, such as the GO navigation system described herein.

DEPR:
Real names, network addresses, and the descriptive information are loaded into the Registry 10 by the Librarian 20. In the preferred embodiment, the Librarian 20 and the Index 30 communicate with the database 12 using an ODBC interface. In the preferred embodiment, the database 12 has a capacity on the order of several hundred million entries. The Registry 10 and database 12 help ensure a consistent structure and vocabulary across Web sites.

DEPR:
The Librarian 20 has a Registration Service 22 and a Crawler 24, each of which is coupled to the database 12 and to a network such as the Internet 50. The Registration Service 22 receives new mappings of real names to network addresses, and descriptive information, and loads them into or "registers" them with the Registry 10. The Registration Service 22 receives the mappings from a client 70 over the Internet 50. The Crawler 24 traverses or crawls the Internet 50, periodically connecting to registered Web servers that are connected to the Internet, to locate changes to the mappings stored in or in association with the Web servers.

DEPR:
A Name File 64 is also stored in association with the Web Server 60 such that the Web Server can retrieve the Name File and forward its contents to the Internet 50 in response to a request. In the preferred embodiment, the Name File 64 stores one or more real name entries. Each real name entry contains a real name of a resource in the Web Server 60, a description of the resource, a network address, or other identifier of the location of the resource, and other information about the resource such as its language and intended geographic region of use. Preferably, the Name File 64 also stores an identifier of a grammar that is used to format the other information in the Name File. In this way, the information in the Name File is self-describing and language-independent.

DEPR:
As indicated by path 29, the Crawler 24 can contact the Web Server 60 and retrieve values stored in the Name File 64 using a connection through the Internet 50. As indicated by path 28, the Crawler 24 can notify the Index 30 that the Index Files 34 need to be updated to reflect a change in the information stored in the Name File 64.

DEPR:
Generally, in the preferred embodiment, the Index Files 34 are more compact than the indexes maintained by conventional search engines, because the amount of information represented in all the Name Files 64 is far less than the total content of all network resources available on the Web. Such compactness is a distinct advantage, providing greater scalability and responsiveness than conventional search engines. In addition, the compact size of the Index Files 34 allows the Index 30 to be replicated in multiple different geographic locations.

DEPR:
The Resolver 40 comprises one or more resolver processes R1, R2, Rn, each of which is coupled respectively to a Service 42, 44, 46. Each resolver process R1, R2, Rn communicates with its respective Service 42, 44, 46 to receive requests containing a real name, convert or resolve the real name into a network address associated with the real name, and forward the network address and other information associated with the real name to the requesting Service.

DEPR:
For example, under control of the browser 74 and the operating system 72, the client 70 can establish an HTTP connection through the Internet 50 to the Registration Service 22. The browser 74 retrieves pages or forms from the Registration Service 22 that are prepared in the HTML language. The browser 74 displays the pages or forms. A user of the client 70 reads the pages, or enters information in a form and sends the filled-in form back to the Registration Service 22. In this way, the client 70 and the Registration Service 22 carry out a dialog by which a user of the client 70 can perform functions offered by the system.

DEPR:
In one embodiment, the system provides a set of customer information management functions that store, track, and update information about customers of the system. The information managed for each customer is called a customer profile. The customer profiles are stored in the database 12.

DEPR:
When the Customer/New Customer option is selected, the system generates one or more Web pages containing forms that enable a user to enter a new customer profile. The form has fields for entry of a name, address, telephone number, contact person, and payment method. The Web pages and forms are communicated to the client 70 and displayed by the browser. The user of the client 70 enters appropriate information into the data entry fields and clicks on or selects a "SUBMIT" button on the Web page. In response, the client 70 returns the filled-in form in an HTTP transaction to the system. The system extracts the entered information from the fields and stores the information in a table of the database 12.

DEPR:
Welcome to the Real Name System registration site. Before you can submit your Real Name addresses, you need to provide us with some information about you and the organization that you may represent.

DEPR:
Preferably, the system then displays a Web page containing a form that enables the system to receive further information about the user. The form has fields for entering the user's name, address, city, state, postal code, nation, and telephone number. The user enters the requested information and clicks on a NEXT button. The system checks each value to verify that it matches the proper data format required for the corresponding field. The values are stored in the database 12 in association with the user's name and email address. Collectively, this information is the customer profile. Once the customer profile is established, the user can create real name entries and store them in one or more Name Files 64.

DEPR:
The primary function offered by the Registration Service 22 is registration of new real names into the Registry 10. In one embodiment, the Registration Service 22 is invoked by selecting the Create option from the top-level menu page. As shown in block 200, an external user or "customer" of the system identifies himself or herself to the system so that information entered later can be associated with the customer. This information includes an electronic mail address of the customer whereby messages can be directed from the Registration Service 22 to the customer over the Internet 50. In this context, the terms "customer" and "user" refer to the operator of a computer remotely connected to the system, for example, the client 70.

DEPR:
As indicated in block 202, the customer then provides information to the Registration

Service 22 that identifies a network resource of the Web Server 60, by its location, its real name, and descriptive information about the network resource. For example, the customer enters the real name "Microsoft Internet Explorer," the URL http://www.microsoft.com/ie4/ aboutie4.html, and a description about the resource. Preferably, this information is entered in fields of a Web page that is constructed for the purpose of receiving the information, in the form shown in Table 3:

DEPR:
When the user has entered all the information, to continue processing of the Name File 64, the user clicks on the NEXT function button at the bottom of the page. In response, as shown in block 204, the Registration Service 22 constructs a Name File 64 based on the information entered by the customer. At this point, the Name File 64 is stored on a server accessible to the Registration Service 22. However, the Name File 64 is not yet stored in association with the Web server 60.

DEPR:
When the user selects the first option ("Live update of a previously registered Name File"), as shown in blocks 214-216, the system activates the Crawler, which locates the user's Name File over the Internet, and updates the database 12, as described below. Thus, the "Live update" function provides a way for a user to force the system to locate a modified Name File and update itself with the new information. Alternatively, as described below in connection with the Crawler, the user may simply wait and the Crawler eventually will locate the modified file and update the database.

DEPR:
When the user selects the second option ("Registration of a new Name File on your website"), as shown in blocks 220 to 222, in response the system constructs and sends to the client 70 a Web page with which the user can enter payment information pertaining to the user and its Name Files. Payment steps of the activation process are an entirely optional part of the process, and other embodiments are contemplated that omit any payment mechanism. In the embodiments that do use a payment mechanism, the Web page contains fields that accept entry of payment information. For example, the fields enable entry of a credit card type, card number, expiration date, and cardholder name. The system receives the payment information values in block 224.

DEPR:
In block 242, the Registration Service 22 notifies the Index Builder 32 that a new entry has been made in the database 12. Path 26 of FIG. 1B represents the notification. The notification includes information sufficient to identify the new entry in the database 12, for example, a row identifier ("rowid") of a table in which the new entry is stored. In response, the Index Builder 32 carries out a live update of the Index Files 34, in the manner discussed further below.

DEPR:
In the preferred embodiment, the database 12 is available to receive queries from registered members of the system. As a result, a registered member can submit queries to the database 12 that request the database to display currently registered information about network resources or Web pages of other organizations. Accordingly, if another registered user succeeds in registering information that misrepresents the content of that user's network resources, the misrepresentation can be reported to the Registry for corrective action. Thus, in this manner, the formality of the registration process, and the open query capability of the database 12 enable the present system to avoid the deception that is possible through the improper use of metatags.

DEPR:
For each of the selected rows or records, in block 304, the Crawler 24 polls the customer Web site that is represented by the row or record, searching for updates to the Name File 64 that is stored in association with that Web site. The polling step includes the steps of opening an HTTP connection to the Web site, requesting and receiving a copy of the Name File. The Crawler 24 parses the Name File, using an XML parser, to identify real name entries, and values within each real name entry, that specify the real name, network address, and descriptive information relating to network resources. An XML parser is commercially available from Microsoft Corporation.

DEPR:
Preferably, the index build requests comprise an identifier, called a FileId, of a file or row that is mapped in the File Info table described above. The Index Builder 32 looks up the FileID in the File Info table and retrieves all entries in the database that match the FileID. Each database entry includes a unique identifier that is associated with a network resource that is described in the database entry. The unique identifiers are generated using a sequence facility of the database server. Based on the unique identifier, for database entry that matches the FileID, the Index Builder retrieves a matching index entry. The information in the index entry is compared to the information in the build request. If the information in the build request is different, the index

entry is updated. If the information in the build request indicates that the associated network resource has become inactive or unavailable in the network, the index entry is deleted.

DEPR:
For example, consider a query for the real name address " Microsoft." Assume that resolution of the query yields no exact match, but yields more than one inexact match, such as "Microsoft Excel" and "Microsoft Word". In the first stage of the ordering process, these two entries would be ranked against relevance criteria and re-ordered if one entry is determined to have greater relevance to the query than the other. The relevance criteria include, for example, the number of words in each entry, whether each entry contains the exact query term, etc. In this example, according to these criteria, each of the two entries has equal relevance; therefore, they are not re-ordered. In the second stage of the ordering process, the Resolver 40 retrieves statistical information about each entry from the Statistics Service described herein. The statistical information includes a usage value for each real name entry that is computed by applying a weighting function to a count of past resolutions for that real name. The weighting function operates to give more weight to recent resolutions for the real name than to resolutions that occurred in the distant past. The Resolver compares the usage values for each of the entries and re-orders the entries, if necessary, so that the entry having the highest-weight usage value is first in order in the Entry Set object.

DEPR:
In block 510, the Resolver 40 formats the response of the index into an output message. In a preferred embodiment, the Resolver 40 constructs an XML file containing the information in the response from the Index 30. In the preferred embodiment, the services 42, 44, 46 each are provided with an XML parser that can convert the XML file produced by the Resolver 40 into text or other information in a format that is usable by the client 70. Also in the preferred embodiment, each entry referenced in the Entry Set object contains a usage value that indicates the number of times that the entry has been resolved. The usage values are used to order the entries when they are displayed or otherwise used by one of the Services 42-46.

DEPR:
In an alternate embodiment, the Resolver 40 is capable of distinguishing among network addresses that refer to resources located on the Internet, an internal business network or "intranet", and an externally accessible internal business network or "extranet". In an intranet environment, the Resolver 40 accesses a Registry 10 that is located within the organization that owns and operates the Resolver. The Registry 10 stores resource information that identifies intranet resources. The Resolver 40 resolves real names entered by the user into the locations of intranet resources, and navigates the user to them.

DEPR:
In an alternate embodiment, when the GO Service 42 is implemented as a browser plug-in installed in the client 70, the GO Service provides character encoding information to the Resolver 40. To obtain the character encoding currently used on the client 70, the GO Service 42 calls an operating system function of the operating system that runs on the client 70. The GO Service 42 attaches the character encoding information to the URL that is used to return the user's query to the Resolver 40. In this way, the Resolver receives information indicating the language and character set currently used by the client 70, and can respond with a network resource that is appropriate to that language.

DEPR:
As described above in connection with the Resolver 40, each time a real name resolution is carried out by the Resolver, it writes a log file entry. The system includes a Statistics Service 82 that is responsible for reading the log file and loading information from the log file into the Index Files 34.

DEPR:
In the preferred embodiment, the Statistics Service 82 operates periodically on scheduled basis. The Statistics Service 82 reads each record of the log file and constructs an index object based on the information in the log file. The Statistics Service 82 then sends a message to the Index Builder 32 that requests the Index Builder to persistently store the values in the Index Files 34. In response, the Index Builder 32 stores the values in the Index Files 34.

DEPR:
When the Statistics & Billing/Statistics option is selected, the system generates a Web page 700 in the form shown in FIG. 7A. The Web page 700 has a list 702 of top-level options. A set of function buttons 704 enable the user to establish other global functions such as resolving an address, entering new customer information, obtaining customer service, and learning more information about the real name system.

DEPR:
The Select Entries button 712 is used to identify a range of entries within a Name File for which statistics are to be generated. When the user selects the Select Entries button 712, the system reads the Name File on the server having an IP address matching the IP address of the user's current domain. The system parses the Name File and displays a list of all the real names in a new Web page that is sent to the client 70. The Web page displays a radio button adjacent to each of the real names in the list. By clicking on the radio button and then submitting the Web page to the system, the system will provide statistical information for all the selected real names in all reports that are generated later.

DEPR:
The Select Time button 714 is used to identify a time frame for which statistics are to be generated. When the user selects the Select Time button 714, the system generates a new Web page and sends it to the client 70. The Web page includes a form into which the user enters a starting date and an ending date. When the user submits the filled-in page to the system, the system receives and stores the date values. When reports are generated thereafter, the reports will contain statistical information for resolutions of real names that occurred within the specified dates.

DEPR:
The Report per Entry button 716 is used to generate a report and graph showing all real name resolutions that have occurred for each real name entry defined in the current Name File. When the Report per Entry button 716 is selected, the system reads statistical information that is stored in the statistical tables of the database 12 for each of the real names that are defined in the current Name File. The system generates a graph and a chart of the statistical information, and generates a Web page containing the graph and chart.

DEPR:
FIG. 7A is an example of a Web page generated in this manner. The graph pane 708 shows an exemplary bar graph. Each bar in the bar graph represents a real name defined in the current Name File. The vertical axis 720 identifies the number (in thousands) of resolutions of each real name. The horizontal axis 722 identifies each name for which statistics information is reported. The statistics pane 710 comprises a real name column 730, a quantity of resolutions column 732, and a percentage column 734. The real name column 730 lists each real name that is defined in the current Name File. The quantity of resolutions column 732 gives the number of resolutions of that real name that have occurred within the currently defined time period. The percentage column 734 indicates, for each real name, the percentage of total resolutions represented by the resolutions of that real name.

DEPR:
In an embodiment, a fee is charged by the owner of the real name system to end users or customers who register real names in the Registry 10. The Librarian 20 records a charge against the account of the user when a new entry is submitted to the system using the Registration Service 22. In another embodiment, end users or customers who register real names in the Registry 10 pay a fee to the owner of the real name system for each resolution executed by the Resolver 40 in response to a third-party request. The Resolver 40 records a charge against the account of the user when each resolution is completed. In these embodiments, the account information and charges are logged and accumulated in tables of the database 12. Periodically, an external billing application reads the charge and account tables of the database 12 and generates invoices that are sent to the user. The Statistics & Billing/Billing Information option of the top-level option list 702 enables the user track and monitor, in real time, the user's credits and payments for registered real name entries, as well as resolution fees. When the Billing Information function is selected, the system reads the charge and account tables of the database 12 and generates a report, in a Web page, summarizing the charges to the customer. The Web page is delivered to the client 70 and displayed by it.

DEPR:
FIG. 8 is a block diagram that illustrates a computer system 800 upon which an embodiment of the invention may be implemented. Computer system 800 includes a bus 802 or other communication mechanism for communicating information, and a processor 804 coupled with bus 802 for processing information. Computer system 800 also includes a main memory 806, such as a random access memory (RAM) or other dynamic storage device, coupled to bus 802 for storing information and instructions to be executed by processor 804. Main memory 806 also may be used for storing temporary variables or other intermediate information during execution of instructions to be executed by processor 804. Computer system 800 further includes a read only memory (ROM) 808 or other static storage device coupled to bus 802 for storing static information and instructions for processor 804. A storage device 810, such as a magnetic disk or optical disk, is provided and coupled to bus 802 for storing information and instructions.

DEPR:
Computer system 800 may be coupled via bus 802 to a display 812, such as a cathode ray tube (CRT), for displaying information to a computer user. An input device 814, including alphanumeric and other keys, is coupled to bus 802 for communicating information and command selections to processor 804. Another type of user input device is cursor control 816, such as a mouse, a trackball, or cursor direction keys for communicating direction information and command selections to processor 804 and for controlling cursor movement on display 812. This input device typically has two degrees of freedom in two axes, a first axis (e.g., x) and a second axis (e.g., y), that allows the device to specify positions in a plane.

DEPR:
Computer system 800 also includes a communication interface 818 coupled to bus 802. Communication interface 818 provides a two-way data communication coupling to a network link 820 that is connected to a local network 822. For example, communication interface 818 may be an integrated services digital network (ISDN) card or a modem to provide a data communication connection to a corresponding type of telephone line. As another example, communication interface 818 may be a local area network (LAN) card to provide a data communication connection to a compatible LAN. Wireless links may also be implemented. In any such implementation, communication interface 818 sends and receives electrical, electromagnetic or optical signals that carry digital data streams representing various types of information.

DEPR:
Network link 820 typically provides data communication through one or more networks to other data devices. For example, network link 820 may provide a connection through local network 822 to a host computer 824 or to data equipment operated by an Internet Service Provider (ISP) 826. ISP 826 in turn provides data communication services through the world wide packet data communication network now commonly referred to as the "Internet" 828. Local network 822 and Internet 828 both use electrical, electromagnetic or optical signals that carry digital data streams. The signals through the various networks and the signals on network link 820 and through communication interface 818, which carry the digital data to and from computer system 800, are exemplary forms of carrier waves transporting the information.

DEPC:
Modifying and Deleting Name File Information

CLPV:
parsing the name file;

CLPV:
retrieving the name file, parsing the name file; building an index entry based on the values parsed from the name file; and storing the index entry in an index of the metadata registry;

ORPL:
William Y. Arms, Christophe Blanchi, Edward A. Overly, D-Lib Magazine, "An Architecture for Information in Digital Libraries,"
http://www.dlib.org/dlib/february97/cnri/02arms1.html, "creation date" Feb., 1997.

WEST

Generate Collection

L9: Entry 9 of 26 File: USPT Aug 8, 2000

DOCUMENT-IDENTIFIER: US 6101491 A
TITLE: Method and apparatus for distributed indexing and retrieval

ABPL:
Systems and methods consistent with the present invention access information from databases having stored indexes of completed information corresponding to the databases by receiving a query identifying desired information; examining the concepts of information in the stored indexes to identify as hits the contents of the databases that match the query; determining, for each hit, a measure of a difference between the query and the conceptual information from the indexes; and combining the hits from the indexes in accordance with the determined measures.

PCPR:
This application is a continuation-in-part under 37 C.F.R. .sctn. 1.60 of U.S. patent application Ser. No. 08/499,268, for "Method and Apparatus for Generating Query Responses in a Computer-Based Document Retrieval System," filed Jul. 7, 1995, now U.S. Pat. No. 5, 724,571, which is incorporated herein by reference.

BSPR:
The present invention relates to text retrieval systems and, more particularly, to a method for distributing indexes containing conceptual information derived from documents and responding to queries using those indexes. The present invention also relates to responding to queries using existing indexes of conventional document retrieval systems by reindexing documents identified by those systems in accordance conceptual information derived from those documents.

BSPR:
There are two main concerns facing text retrieval systems: (1) How to identify terms in documents that should be included in the index; and (2) After indexing the terms, how to determine that a document matches a query? Conventional text retrieval techniques rely on indexing keywords in documents. Index terms can be from single words, noun phrases, and subject identifiers derived from syntactic and semantic analysis. Conventional text retrieval systems for the World Wide Web, such as Yahoo!.TM. from Yahoo! Inc. and AltaVista.TM. from Digital Equipment Corporation, use these and other types of keyword indexing techniques to index documents available on the web. Unfortunately, a document's keywords alone rarely capture the document's true contents. Consequently, systems relying on keywords in an index to retrieve documents in response to queries often provide unsatisfactory retrieval performance.

BSPR:
Yahoo!, AltaVista, and other convention text retrieval systems for the web employ programs called "web crawlers" to traverse the web. Web crawlers follow links from page to page and extract terms from all the pages that they encounter. Each search engine then makes the resulting information accessible by providing lists of specific pages that match an input search request or query.

BSPR:
Because the web constantly changes as existing pages are modified and new pages are added, web crawlers cannot simply traverse the web and index it once. Instead, to stay current, they must repeatedly traverse the web to

identify changes for refreshing the index. Changes are made constantly and without notice, however, so it is not possible to keep up with them.

BSPR:
Moreover, many sites on the web are now reluctant to provide the access demanded by web crawlers to access and index the sites pages because the resources given to the web crawler detract from those for the users. This poses another problem to the ongoing success of such retrieval techniques on the web.

BSPR:
"WAIS," which stands for Wide Area Information Servers, suggests one alternative to the use of web crawlers for indexing. WAIS is an architecture for a distributed information retrieval system based on the client server model of computation. WAIS allows users of computers to share information using a common computer-to-computer protocol. WAIS was originally designed and implemented by a development team at Thinking Machines, Inc. led by Brewster Kahle. WAIS requires the sites that publish information on the web to publish an index of that information as well. Search engines can then use the published indexes to respond to queries. Although WAIS helps the resource problem associated with web crawler-based text retrieval systems, it fails to address a more fundamental problem with conventional search and retrieval systems: the quality of the ranked output.

BSPR:
The quality of the output suffers from the way most searches occur. The most common methods for determining whether a document matches a query are the "boolean model" and the "statistical model." According to the boolean model, a match occurs when a document's index terms meet the boolean expression given by the user. The statistical model, on the other hand, is based on the similarity between statistical properties of the document and the query.

BSPR:
It is not unusual for conventional search engines using either approach to return a large number of matches for a simple query. When faced with a list of 20,000 hits in response to a query--not an uncommon experience when searching the web--a user cannot effectively review all the results. Whether the user accesses the matches serially or randomly, the review process takes an unwieldy amount of time to locate the documents of particular interest. Typically, Internet web searchers provide the user with the first 10 hits and continue to provide additional blocks of 10 until the user finds something acceptable or gives up. If the user has a simple information need and the answer shows up in the first 10 or 20 hits, then this is not unreasonable. However, if the user has serious research interest in the results, then it may be important to see the information available in the remaining hits.

BSPR:
Consequently, the criteria by which these hits are ranked becomes very important. More and more systems support some type of ranking feature because users have demanded easy-to-use query languages and ranking to sort out the most important information.

BSPR:
WAIS supports one document ranking scheme. WAIS scores documents based on the number of occurrences of a query term in a document, the location of the terms in a document, the frequency of those terms within the collection, and the size of the document. WAIS, however, uses a least-common-denominator standard that does not allow for sophisticated querying and ranking of results.

BSPR:
At the same time, the growing volume of material for indexing has required search engine designers to focus on techniques for efficiency and volume processing, rather then on techniques for guaranteeing the best possible rankings. The conflict between these two objectives, accurate search results and indexing huge collections of information, poses a significant problem for the developers of the next generation of text retrieval systems.

BSPR:
Accordingly, systems and methods consistent with the present invention

substantially obviate one or more of the problems due to limitations, shortcomings, and disadvantages of the related art by distributing the process of indexing documents using the conceptual indexing approach among multiple processes or platforms, applying queries to each index individually, and combining the results using penalty-based scores that include a measure of the difference between terms of the query and the conceptual terms found in the index. A method consistent with the present invention for accessing information from databases having stored indexes of completed information corresponding to the databases comprises the steps, performed by a processor, of: receiving a query identifying derived information; examining the concepts of information in the stored indexes to identify as hits the contents of the databases that match the query; determining, for each hit, a measure of a difference between the query and the conceptual information from the indexes; and combining the hits from the indexes in accordance with the determined measures.

DRPR:
FIG. 4 is a flow chart of the steps performed by a query server consistent with the present invention;

DRPR:
FIG. 5 is a flow chart of the steps performed by a query dispatcher/aggregator of the distributed text retrieval system consistent with the present invention;

DEPR:
Conceptual indexing refers to extracting conceptual phrases from the source material, assimilating them into a hierarchically-organized, conceptual taxonomy, and indexing those concepts in addition to indexing the individual words of the source text. Dynamic passage retrieval refers to a technique for using the positional information about where words and concepts occur in text to locate specific passages of material within the text that are responsive to a query.

DEPR:
information resides. The processes may be located on a single machine or on multiple platforms in one or more networks. Thus, the bulk of the maintenance of the indexes is done by the information providers rather than by centralized text retrieval systems. This eliminates the need for sites to provide service for repeated requests by programs such as web crawlers that traverse their pages to see if anything has changed. Rather, the sites perform their own indexing and provide a service to retrieval requests.

DEPR:
In this scheme, the sites employ a "push" model rather than a "pull" model for indexing. Rather than waiting for central indexers to pull that information from the site by repeated polling, the sites know when a page has changed and incrementally update their local index. Central to this architecture is an attribute of the dynamic passage retrieval algorithm that enables result lists from independent searches to be easily combined. Because the penalty scores assigned to passages by the relaxation-ranking algorithm are independent of collection size or statistics, the results of queries to different sites can be collated together and pruned on the basis of their penalty scores, without risk of losing more important information in favor of less important information.

DEPR:
FIG. 1 illustrates the components of a distributed indexing and retrieval system 110 consistent with the present invention. System 110 includes a user application 120, a query dispatcher/aggregator 130, and multiple index managers 140a and 140b. Although system 110 in FIG. 1 includes two index managers, more than two may be used to take full advantage of the principles of the present invention.

DEPR:
System 10 resides either in a single platform, such as a personal computer, workstation, or mainframe, or in a network, such as the Internet or an Intranet. System 10 may also be partitioned among multiple processes or platforms. For example, user application 120 may reside on a platform

different from the platforms for query dispatcher/aggregator 130 and index manager 140.

DEPR:
Thus, the content or material to be indexed is generally partitioned into separate domains, each managed by an index manager (140a or 140b). Index manager 140a or 140b is either specially configured to include functionality like that described below with reference to FIGS. 2-4, or configured to include functionality to integrate the system with legacy document retrieval systems like Yahoo! and AltaVista (see FIGS. 4, 6-7). Alternatively, an index manager itself can be configured as a query dispatcher/aggregator, integrating other index managers in a manner similar to way query dispatcher/aggregator 130 integrates index managers 140 in FIG. 1.

DEPR:
User application 120, for example, a web browser such as Netscape or Internet Explorer, receives user queries, including a term or combination of terms, and a set of parameters, and passes them to dispatcher/aggregator 130. This process uses a protocol for communicating queries and results between user application 120 and query dispatcher/aggregator 130, for example, the TCP/IP protocol used in the Internet. User application 120 receives the query terms from the user and the parameters from predetermined tables that may be modified by user preferences, and sends the query and parameters to query dispatcher/aggregator 130. (In an alternative configuration, the server upon which query dispatcher/aggregator 130 resides, provides user application 120 with a web page to enter the query and search parameters. After the user enters this data, user application 120 sends it to query dispatcher/aggregator 130 using the TCP/IP protocol.)

DEPR:
The parameters assist in the process of selecting and scoring hits. One typical parameter specifies the maximum number of hits desired (i.e., a hit limit parameter). Alternatively, query dispatcher/aggregator 130 uses a predetermined hit limit. Other parameters set criteria used in identifying hits from the conceptual index and determining penalty scores for the hits in accordance with user preferences. For example, a parameter may govern the value of a penalty score for things like missing terms from the hit.

DEPR:
Query dispatcher/aggregator 130 passes the query to index managers 140a and 140b, and collects and aggregates the results, including hits and corresponding scores. The hits are either identifiers for documents or passages within the documents, the documents themselves, or the passages within the documents that most closely match the input query. The scores are generated using the penalty-based algorithm that assigns a score based on a measure of the difference between a passage in the document and the query.

DEPR:
Query dispatcher/aggregator 130 collects the hits from index managers 140a and 140b in accordance with a specified hit limit parameter and returns scored hits together with their penalty scores to user application 120. Query dispatcher/aggregator 130 also uses the penalty scores assigned to the hits by the individual index managers 140a and 140b to collate the results into a merged list in increasing order of penalty, preferably eliminating duplicates if they are encountered. The hits with the highest penalty scores are pruned, if necessary, to reduce the resulting aggregated list to the maximum number of hits requested.

DEPR:
Asynchronously, and independently from the query processing, index managers 140a and 140b for the different partitions update their indexes according to the policies of their host sites, for example, web servers holding the content from which the index is built. Host site policies are based, for example, on a calender-driven process such as processing the index overnight or on a push model in which the index is updated whenever a site specific application notifies it of a page that needs to be indexed or reindexed. Thus, index managers 140 update the indexes dynamically and in real time, so they remain as current as the publishing host site chooses.

DEPR:
Index manager 210 has two main functions: (1) building or modifying index 230, and (2) responding to queries from dispatcher/aggregator 130. These functions are performed by index server 240 and query server 220, respectively.

DEPR:
Query server 220 processes incoming document retrieval requests from query dispatcher/aggregator 130. Each request includes a query with parameters. If query dispatcher/aggregator 130 does not provide a hit limit parameter, query server 220 uses its own predetermined hit limit when processing requests. The predetermined hit limit may simply be the number of the hits, the number of the hits with penalty scores that do not exceed a particular value, or all hits regardless of the penalty scores, provided there is some correspondence between the query and the document or passage (e.g., paragraph or relevant section) within the document.

DEPR:
Query server 220 accesses conceptual index 230 to identify matches for the query, i.e., hits, and assigns scores to the hits in accordance with the penalty-based scoring algorithm. Query server 220 then returns the hits and scores to query dispatcher/aggregator 130 in accordance with the hit limit.

DEPR:
The taxonomy can be used as an aid in both formulating and processing queries. In querying the index, terms are treated as concepts and are expanded by their specific children in the taxonomy. Likewise, the taxonomy places limitations on the range of concepts that may correspond to query terms. For additional information and examples of conceptual indexing, see U.S. patent application Ser. No. 08/797,630, entitled, "Intelligent Network Browser Using Incremental Conceptual Indexer," filed Feb. 7, 1997.

DEPR:
Query Server

DEPR:
FIG. 4 is a flow chart of the steps performed by query server 220. First, query server 220 receives a query and parameters from query dispatcher/aggregator 130 (step 410). The hit limit parameter may be one set by user application 120 who submitted the query, by query dispatcher/aggregator 130, or by query server 220.

DEPR:
Query server 220 then accesses index 230 to identify documents or passages in documents corresponding to conceptual information in index 230 that most closely correspond to the query (step 420). Query server 220 scores these hits using a scoring algorithm that scores passages by measuring how much they depart (in any of several dimensions) from an ideal passage, i.e., an exact replica of the query (step 430). The measure is referred to as relaxation ranking. In contrast with traditional retrieval ranking methods, where scores of results are based on accruing weights corresponding to pieces of evidence that a given result is relevant to a query, the scores assigned by the relaxation-ranking algorithm are based on accruing penalties for various kinds of departure from the ideal. Thus, the best passage is the one with the lowest score, as opposed to the highest score used by customary approaches. This approach is referred to as penalty-based scoring.

DEPR:
the additional attractive property that the values of the scores themselves are meaningful and interpretable. Thus, a user looking at a score can determine whether a match is likely to be good or not and can estimate how good it is likely to be. For example, zero (0) is a perfect score and many retrieved results will achieve this score. In contrast, scores assigned by traditional methods are only relatively comparable, and then only when derived from the same collection. Even in the case of probabilistic retrieval, where the scores are estimates of probabilities of relevance, and therefore should be somewhat interpretable, the individual probability scores are relative to the statistics of the collection and not individually meaningful. The probability of one (1) is virtually never reached, and there is no a priori probability that corresponds to a definitely relevant match. The

aforementioned patent application, Ser. No. 08/499,268, for "Method and Apparatus for Generating Query Responses in a Computer-Based Document Retrieval System," describes penalty-based scoring in greater detail.

DEPR:
Returning to FIG. 4, query server 220 returns the hits and corresponding scores to query dispatcher/aggregator 130 (step 440).

DEPR:
Query Dispatcher/Aggregator

DEPR:
FIG. 5 is a flow chart of the steps performed by query dispatcher/aggregator 130. First, query dispatcher/aggregator 130 receives a query and parameters from, for example, user application 120 (step 510). Query dispatcher/aggregator 130 then passes the query to each of the distributed index managers 140 and, particularly, the query server 230 of each index manager 140 (step 520). After each query server 230 processes the query using the associated conceptual index 230, query dispatcher/aggregator 130 receives the hits and scores (step 530). Query dispatcher/aggregator 130 then merges the hits and scores from the various index managers 140 (step 540) and prunes the results to, for example, eliminate duplicates or hits with scores above a threshold value (step 550). Finally, query dispatcher/aggregator 130 returns the results, including the hits and scores to user application 120 (step 560).

DEPR:
Cascaded indexing and retrieval involves the dynamic construction of a conceptual index of information identified by the results of a conventional text retrieval system such as Yahoo! and AltaVista. In order to provide for material that is already indexed by some other methodology that does not provide commensurate penalty-based scores, for example, the methodology used by AltaVista a reindexer takes the results of the conventional search and reindexes the documents, such as web pages, using the relaxation ranking method of the dynamic passage retrieval algorithm. The reindexer then provides the results of this reindexing process to the query dispatcher/aggregator. The reindexer interacts with the conventional index server of AltaVista, passing the query to that server. The reindexer then indexes the contents of the documents identified by the server in response to the query.

DEPR:
FIG. 6 illustrates the components of an index manager 610 consistent with the present invention for implementing cascaded indexing and retrieval. Index manager 610 consists of query server 220, index 230, and a reindexer 620. Index manager 610 is designed to complement a conventional document retrieval system 625, which consists of a document retrieval server 630 and an index 650 of content 640, such as web pages. In the Internet, users send queries to document retrieval system 625 using the TCP/IP protocol, and system 625 in turn accesses index 650 to identify specific web pages that satisfy the terms of each query according to predetermined criteria set by system 625.

DEPR:
To implement cascaded indexing in a manner consistent with the present invention, query server 220 provides the user's query to reindexer 620. Reindexer 620 formats the query for system 630 and transmits the reformatted query to server 630. Server 630 provides the query results to reindexer 620, which accesses the content identified in the hits and reindexes the content into conceptual index 230. Query server 220 then processes the query on index 230 in the manner discussed above, and returns the hits and scores to user application 120.

DEPR:
Cascaded indexing uses the query server processing discussed above with reference to FIG. 4, with the additional step of providing the query to reindexer 620 before accessing index 230 to process the input query.

DEPR:
FIG. 7 is a flow chart of the steps performed by reindexer 620. First, reindexer 620 receives the query from query server 220 (step 710), reformats

the query for document retrieval system 625 (step 720), and then sends the reformatted query to server 630 (step 730). After reindexer 620 sends the query, it receives the hits of document retrieval system 625 (step 740). Reindexer 620 reindexes the documents related to the hits identified by system 625 into index 230 (step 750).

DEPL:
In this representation, the "computer" concept is a more general form of the "laptop" concept. Thus, the "computer" concept is depicted as a parent of the "laptop" concept in the graph structure. The taxonomy can be used alone to organize information for browsing, or it can be used as an adjunct to search and retrieval techniques to construct improve query results.

CLPV:
receiving a query identifying desired information;

CLPV:
distributing the query to the indexes;

CLPV:
examining the concepts of information in the stored indexes to identify as hits the contents of the databases that match the query;

CLPV:
determining, for each hit, a measure of a difference between the query and the conceptual information from the indexes; and

CLPV:
ranking the hits in accordance with the measure for each hit, the hits with lower measures indicating a better correspondence between the represented conceptual information and the query than the hits with higher measures.

CLPV:
identifying concepts from the stored taxonomy that correspond to the query.

CLPV:
identifying concepts from the stored taxonomy that correspond to the query based on the relationships among the concepts in the taxonomy.

CLPV:
receiving a query term and a search parameter setting user preference for text retrieval.

CLPV:
computing the measure of a difference between the query and the conceptual information from one of the indexes corresponding to each hit in accordance with the search parameter.

CLPV:
receiving means configured to receive a query identifying desired information;

CLPV:
distributing means configured to distribute the query to the indexes;

CLPV:
examining means configured to examine the concepts of information in the stored indexes to identify as hits the contents of the databases that match the query;

CLPV:
determining means configured to determine, for each hit, a measure of a difference between the query and the conceptual information from the indexes; and

CLPV:
ranking means configured to rank the hits in accordance with the measure for each hit, the hits with lower measures indicating a better correspondence between the represented conceptual information and the query than the hits

with higher measures.

CLPV:
identifying means configured to identify concepts from the stored taxonomy that correspond to the query.

CLPV:
identifying means configured to identify concepts from the stored taxonomy that correspond to the query based on the relationships among the concepts in the taxonomy.

CLPV:
means configured to receive a query term and a search parameter setting user preference for text retrieval.

CLPV:
computing means configured to compute the measure of a difference between the query and the conceptual information from one of the indexes corresponding to each hit in accordance with the search parameter.

CLPV:
a receiving module configured to receive a query identifying desired information;

CLPV:
a distribution module configured to distribute the query to the indexes;

CLPV:
an examining module configured to examine the concepts of information in the stored indexes to identify as hits the contents of the databases that match the query;

CLPV:
a determining module configured to determine, for each hit, a measure of a difference between the query and the conceptual information from the indexes; and

CLPV:
a ranking module configured to rank the hits in accordance with the measure for each hit, the hits with lower measures indicating a better correspondence between the represented conceptual information and the query than the hits with higher measures.

CLPV:
an identifying module configured to identify concepts from the stored taxonomy that correspond to the query.

CLPV:
an identifying module configured to identify concepts from the stored taxonomy that correspond to the query based on the relationships among the concepts in the taxonomy.

CLPV:
a module configured to receive a query term and a search parameter setting user preference for text retrieval.

CLPV:
a computing module configured to compute the measure of a difference between the query and the conceptual information from one of the indexes corresponding to each hit in accordance with the search parameter.

CLPV:
receiving a query identifying desired information;

CLPV:
distributing the query to the indexes;

CLPV:
identifying information corresponding to the query from the stored indexes

distributed among the plurality of platforms; and

CLPV:
determining a measure of a difference between the query and the identified information.

CLPV:
ranking the identified information from all of the indexes, with the information having lower measures indicating a better correspondence between that information and the query than the information with higher measures.

CLPV:
receiving means configured to receive a query identifying desired information;

CLPV:
distributing means configured to distribute the query to the indexes;

CLPV:
identifying means configured to identify information corresponding to the query from the stored indexes distributed among the plurality of platforms; and

CLPV:
determining means configured to determine a measure of a difference between the query and the identified information.

CLPV:
ranking means configured to rank the identified information from all of the indexes, with the information having lower measures indicating a better correspondence between that information and the query than the information with higher measures.

CLPV:
a receiving module configured to receive a query identifying desired information;

CLPV:
a distributing module configured to distribute the query to the indexes;

CLPV:
an identifying module configured to identify information corresponding to the query from the stored indexes distributed among the plurality of platforms; and

CLPV:
determining means configured to determine a measure of a difference between the query and the identified information.

CLPV:
correspondence between that information and the query than the information with higher measures.

CLPV:
receiving a query identifying desired information;

CLPV:
distributing the query to the indexes;

CLPV:
identifying information corresponding to the query from the stored indexes distributed among the plurality of processes; and

CLPV:
determining a measure of a difference between the query and the identified information.

CLPV:
ranking the identified information from all of the indexes with the

- information having lower measures indicating a better correspondence between that information and the query than the information with higher measures.

CLPV:
receiving means configured to receive a query identifying desired information;

CLPV:
distributing means configured to distribute the query to the indexes;

CLPV:
identifying means configured to identify information corresponding to the query from the stored indexes distributed among the plurality of processes; and

CLPV:
determining means configured to determine a measure of a difference between the query and the identified information.

CLPV:
ranking means configured to rank the identified information from all of the indexes with the information having lower measures indicating a better correspondence between that information and the query than the information with higher measures.

CLPV:
a receiving module configured to receive a query identifying desired information;

CLPV:
a distributing module configured to distribute the query to the indexes;

CLPV:
an identifying module configured to identify information corresponding to the query from the stored indexes distributed among the plurality of processes; and

CLPV:
a determining module configured to determine a measure of a difference between the query and the identified information.

CLPV:
a ranking module configured to rank the identified information from all of the indexes with the information having lower measures indicating a better correspondence between that information and the query than the information with higher measures.

WEST

Generate Collection

L9: Entry 12 of 26 File: USPT Jul 25, 2000

US-PAT-NO: 6094649
DOCUMENT-IDENTIFIER: US 6094649 A
TITLE: Keyword searches of structured databases
DATE-ISSUED: July 25, 2000

US-CL-CURRENT: 707/3; 707/4, 707/5

APPL-NO: 8/ 995700
DATE FILED: December 22, 1997

WEST

Generate Collection

L9: Entry 14 of 26 File: USPT Jun 20, 2000

US-PAT-NO: 6078914
DOCUMENT-IDENTIFIER: US 6078914 A
TITLE: Natural language meta-search system and method
DATE-ISSUED: June 20, 2000

US-CL-CURRENT: 707/3; 707/2, 707/4, 707/5

APPL-NO: 8/ 769929
DATE FILED: December 9, 1996

WEST

Generate Collection

L9: Entry 14 of 26 File: USPT Jun 20, 2000

US-PAT-NO: 6078914
DOCUMENT-IDENTIFIER: US 6078914 A

TITLE: Natural language meta-search system and method

DATE-ISSUED: June 20, 2000

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Redfern; Darren M.	Stratford	N/A	N/A	CAX

ASSIGNEE-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY	TYPE CODE
Open Text Corporation	N/A	N/A	N/A	CAX	03

APPL-NO: 8/ 769929
DATE FILED: December 9, 1996

INT-CL: [7] G06F 17/30
US-CL-ISSUED: 707/3; 707/2, 707/4, 707/5
US-CL-CURRENT: 707/3; 707/2, 707/4, 707/5
FIELD-OF-SEARCH: 707/2, 707/3, 707/4, 707/5, 707/1

PRIOR-ART-DISCLOSED:

U.S. PATENT DOCUMENTS

Search Selected | Search ALL

	PAT-NO	ISSUE-DATE	PATENTEE-NAME	US-CL
☐	5278980	January 1994	Pedersen et al.	707/4
☐	5500920	March 1996	Kupiec	395/2.79
☐	5576954	November 1996	Driscoll	707/3
☐	5640553	June 1997	Schultz	707/5
☐	5642502	June 1997	Driscoll	707/5

FOREIGN PATENT DOCUMENTS

FOREIGN-PAT-NO	PUBN-DATE	COUNTRY	US-CL
0597630	May 1994	EPX	
0638870	February 1995	EPX	
2296799	July 1996	GBX	
90/08360	July 1990	WOX	
WO 95/12173	May 1995	WOX	

OTHER PUBLICATIONS

"Navigating With A Web Compass", R. Baldazo, vol. 21, No. 3, Mar. 1, 1996.
"Alephweb: A Search Engine Based On The Federated Structure", Proc. 7th Joint European Networking Conference, May 13-16, 1996, Bupadest.
DataBase Inspec Instruction of Electrical Engineers, Stevenage, GB, see Abstract, Dec. 1996.

ART-UNIT: 271
PRIMARY-EXAMINER: Black; Thomas G.
ASSISTANT-EXAMINER: Min; Donald
ATTY-AGENT-FIRM: Fitzpatrick, Cella, Harper & Scinto

ABSTRACT:

A meta search system accepts natural language queries which are parsed to extract relevant content, this relevant content being formed into queries suitable for each of a selected number of search engines and being transmitted thereto. The results from the search engines are received and examined and a selected number of the information sources represented therein are obtained. These obtained information sources are then examined to rank their relevance to the extracted relevant content and the portions of interest in each of these ranked information sources are determined. The determined portions are output to the user in ranked order, having first been processed to clean up the portions to include valid formatting and complete paragraphs and/or sentences.

13 Claims, 17 Drawing figures